scientific report

A domesticated transposon mediates the effects of a single-nucleotide polymorphism responsible for enhanced muscle growth

Falk Butter[1], Dennis Kappei[2], Frank Buchholz[2], Michiel Vermeulen[1†+] & Matthias Mann[1++]

[1]Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried, Germany, and [2]Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany

Single-nucleotide polymorphisms (SNPs) in the regulatory regions of the genome can have a profound impact on phenotype. The G3072A polymorphism in intron 3 of insulin-like growth factor 2 (*IGF2*) is implicated in higher muscle content and reduced fat in European pigs and is bound by a putative repressor. Here, we identify this repressor—which we call muscle growth regulator (MGR)—by using a DNA protein interaction screen based on quantitative mass spectrometry. MGR has a bipartite nuclear localization signal, two BED-type zinc fingers and is highly conserved between placental mammals. Surprisingly, the gene is located in an intron and belongs to the hobo-*Ac*-Tam3 transposase superfamily, suggesting regulatory use of a formerly parasitic element. In transactivation assays, MGR differentially represses the expression of the two SNP variants. Knockdown of MGR in C2C12 myoblast cells upregulates Igf2 expression and mild overexpression retards growth. Thus, MGR is the repressor responsible for enhanced muscle growth in the *IGF2* G3072A polymorphism in commercially bred pigs.

INTRODUCTION

In recent years, mass spectrometry (MS)-based proteomics has become a method of choice to characterize protein interactions.

[1]Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Am Klopferspitz 18, Martinsried 82152, Germany
[2]Max Planck Institute of Molecular Cell Biology and Genetics, Pfotenhauerstrasse 108, Dresden 01307, Germany
†Present address: Department of Physiological Chemistry and Cancer Genomics Centre, University Medical Center Utrecht, 3584 CG Utrecht, The Netherlands
+Corresponding author. Tel: +31 88 75 508 29; Fax: +31 88 75 681 01;
E-mail: m.vermeulen-3@umcutrecht.nl
++Corresponding author. Tel: +49 89 8578 2558; Fax: +49 89 8578 2219;
E-mail: mmann@biochem.mpg.de

In a quantitative and high-resolution format it enables proteins binding specifically to various types of bait molecules to be distinguished from background binders (Vermeulen *et al*, 2008). We have previously used stable isotope labelling with amino acids in cell culture (SILAC; Ong *et al*, 2002) to determine the binding of transcription factors to specific DNA elements (Mittler *et al*, 2009). This method is generic and suited to identifying novel regulatory proteins binding differentially to DNA sequences with single-nucleotide resolution.

Previously, it has been reported that the G3072A single-nucleotide polymorphisms (SNP) in the imprinted insulin-like growth factor 2 (*IGF2*) locus in pigs is associated with less body fat and increased muscle mass. The Q allele (A-SNP) seems to abrogate the binding of a putative repressor (van Laere *et al*, 2003). This SNP association has been verified independently (Jungerius *et al*, 2004; Estelle *et al*, 2005; Oczkowicz *et al*, 2009) and differential binding of this putative repressor protein to the SNP might be connected directly with regulation of muscle growth. Using our quantitative MS-based approach, we aimed to identify this putative repressor.

RESULTS

Screen for differential *IGF2* G3072A interactors

We performed DNA pulldowns with nuclear extracts of a SILAC-labelled mouse myoblast cell line (C2C12) to discover the repressor that binds differentially to the G3072A SNP in pigs. In this screen, the differential abundance of proteins bound to biotin-immobilized oligonucleotides containing either the q (G variant) or Q (A variant) appears as a ratio between the light and heavy peptides (Fig 1A). Among more than 1,000 proteins that were identified and quantified, Uniprot entry q3umd3 had the largest SILAC ratio (5.5). In an independent experiment, q3umd3 again had the highest ratio and it consistently had the lowest ratio after switching SILAC labels (Fig 1B). In a two-dimensional interaction plot of the quantitative results, q3umd3 is seen to be the only
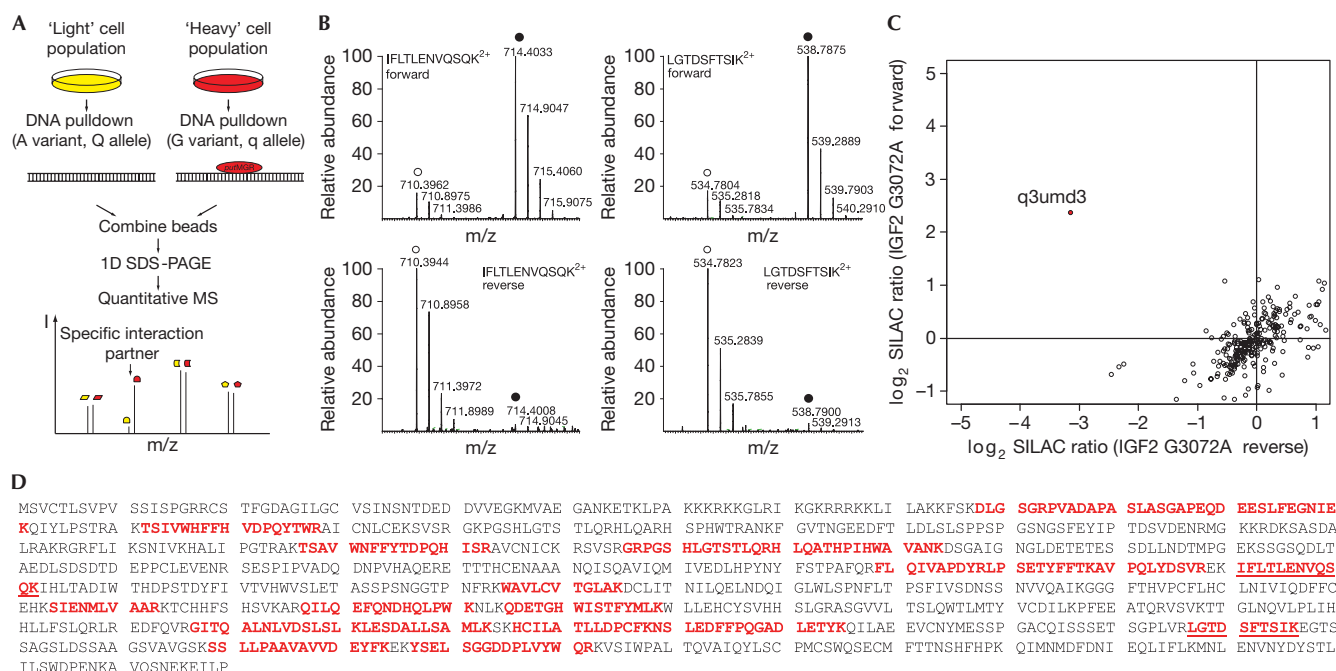
Fig 1 | Identification of muscle growth regulator. (A) A schematic of the quantitative-SILAC-based DNA interaction screen with DNA oligonucleotides containing either the q variant or Q variant of the SNP. Specific interaction partners are differentiated from background binders by having a SILAC ratio other than 1:1. **(B)** The mass spectrum of two representative peptides of MGR detected in the quantitative MS screen for differential interaction partners. In the forward pulldown, the heavy form of the peptides IFLTLENVQSQK^{2+} and LGTDSFTSIK^{2+} are enriched over the light form (q variant of the SNP incubated with heavy extract, filled circle), whereas in the reverse pulldown the light form of the peptide is enriched (q variant of the SNP incubated with light extract, open circle). **(C)** Two-dimensional interaction plot for all detected proteins shows specific enrichment of a single protein (Uniprot entry q3umd3) with the q-probe against the Q-probe. **(D)** The complete open reading frame of the MGR gene, with identified peptides from the three pulldowns (bold red). MGR, muscle growth regulator; MS, mass spectrometry; SDS-PAGE, SDS-polyacrylamide gel electrophoresis; SILAC, stable isotope labelling with amino acids in cell culture; SNP, single-nucleotide polymorphism.

protein with a high forward and low reverse ratio (Fig 1C). None of the other binders had consistent and significant ratios in the three experiments. This showed that q3umd3 binds specifically to the q allele and made it the prime candidate resulting from our proteome-wide, quantitative MS-based screen for the putative repressor.

The putative repressor is a transposable element

We identified the putative repressor with 114 tryptic peptides measured with 450 ppb average absolute mass deviation and an average Mascot search score of 38. In total, 19 sequence-unique tryptic peptides, which cover 31% of the coding sequence and map the corresponding genomic location with high confidence, were identified (Fig 1D; supplementary Table S1 online). Interestingly, this region was not annotated as gene coding in the Ensembl genome browser (www.ensembl.org), but as a 7.2 kb intron of a zinc-finger protein of unknown function, *ZC3H11A* (Fig 2A). Analysis of the intron showed an open reading frame (ORF) of 2,940 nucleotides (in mice) with an ATG start codon and a TAA stop codon that is highly conserved among placental mammalian genomes. The GERP (Genomic Evolutionary Rate Profiling) score—which is a measure for phylogenetic conservation (Cooper *et al*, 2005)—of the ORF is comparable with that of the exonic regions of other genes. Various expressed sequence tag

clones cover the entire sequence of the ORF; together with our mass spectrometric data, this clearly shows that the region is transcribed and translated actively. Notably, a high CoreBoost promoter score (Zhao *et al*, 2007) for the adjacent region 5′ of the ORF suggests that transcriptional control for this ORF might be independent of the surrounding gene. Given its putative role in regulating muscle growth, we name this protein muscle growth regulator (MGR).

The MGR gene encodes a 110 kDa protein with a distinct domain architecture (Fig 2B), including a classical bipartite nuclear localization signal as well as two BED-type zinc fingers, typical DNA interaction domains of chromatin-boundary-element-binding proteins and transposons (Aravind, 2000). A hobo-*Ac*-Tam3 (hAT) dimerization domain clearly places this non-annotated gene in the hAT transposable element superfamily (Kempken & Windhofer, 2001; Rubin *et al*, 2001). This transposon family is ancient and exists in fungi, plants and animals, and there are examples of domesticated hAT transposons that function as developmental regulators in their hosts (Feschotte & Pritham, 2007).

MGR binds differentially to the *IGF2* polymorphism

To further characterize MGR, we recombinantly expressed the longest murine ORF of 980 amino acids with a glutathione-*S*-transferase tag in *Escherichia coli* and tested its binding to the
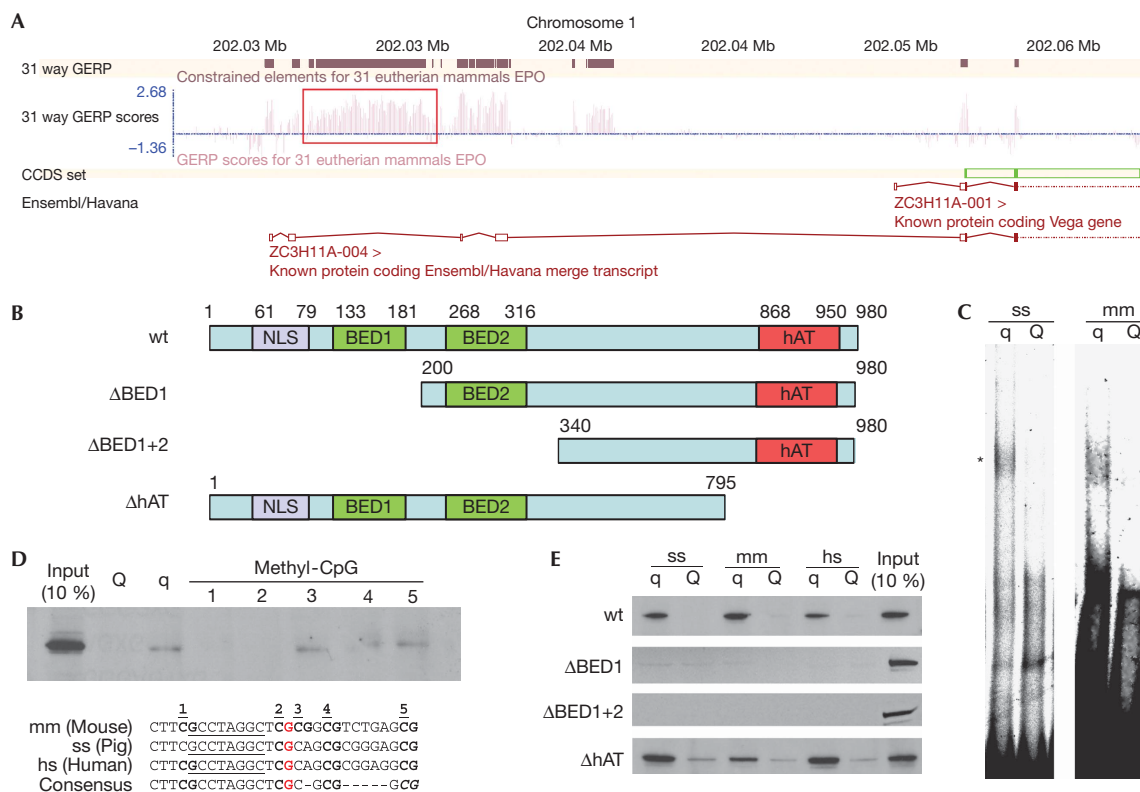
Fig 2 | Characterization of muscle growth regulator. (**A**) A schematic of the genomic region containing the MGR from Ensembl. MGR is a single-exon gene (red box) contained in the intronic region of *ZC3H11A*. This region shows high conservation (GERP) scores in placental mammals. (**B**) A schematic of MGR (wt) and the constructed deletion variants. MGR contains four annotated domains: a classical nuclear localization signal (NLS), two BED-type zinc fingers and a hAT domain. (**C**) The recombinant murine MGR (asterisk indicates the protein–DNA complex) shifts with the q variant of the SNP, but not with the Q variant at the porcine and murine Igf2 sequence. (**D**) Methylation-dependent binding is observed for the two methyl-CpG islands directly adjacent to the palindromic sequence. (**E**) The murine full-length and ΔhAT MGR bind to the q variant, but not the Q variant of mouse, pig and human sequence in DNA pulldown experiments with recombinant GFP–MGR. Deletion variants of the first or both BED fingers are not able to bind to any Igf2 sequence. GERP, Genomic Evolutionary Rate Profiling; GFP, green fluorescent protein; hs, *Homo sapiens*; MGR, muscle growth regulator; mm, *Mus musculus*; SNP, single-nucleotide polymorphism; ss, *Sus scrofa*; wt, wild type.

IGF2 DNA sequence containing the SNP by electrophoretic mobility-shift assay. The recombinant protein shifted specifically with the q allele of the porcine *IGF2* sequence, confirming the mass spectrometric identification of MGR (Fig 2C). In addition, MGR also shifted the oligonucleotide containing the murine Igf2 locus, showing that a conserved region in the oligonucleotide sequence is responsible for binding. There was no detectable interaction between MGR and any of the Q variants. The binding of the putative repressor to the q allele is CpG methylation-dependent (van Laere *et al*, 2003). Consistently, we observed loss of interaction between MGR and the q allele when the CpG pairs directly adjacent to the palindromic sequence were methylated (Fig 2D). This further indicates that the palindromic sequence in the oligonucleotide represents the MGR DNA-binding motif. To investigate the importance of the zinc fingers for MGR DNA-binding activity, we constructed deletion variants lacking the first (ΔBED1) or both DNA-binding domains (ΔBED1 + 2; Fig 2B). As expected, deletion of one or both zinc fingers abrogates DNA binding (Fig 2B). A third mutant that lacks the carboxy-terminal dimerization domain still shows differential DNA-binding activity (Fig 2E).

MGR functions as a repressor at the Igf2 promoter

To investigate the effect of MGR on Igf2 expression, we established transactivation assays using the sequence of the intronic murine Igf2 DNA sequence containing the q variant or Q variant fused to the endogenous murine Igf2 P3 promoter followed by the luciferase reporter gene. The q-variant-containing construct showed significantly lower luciferase activity compared with the Q variant in C2C12 cells, which is consistent with previous findings (van Laere *et al*, 2003). To test whether MGR is a repressor of Igf2 expression, we investigated a potential differential effect of these two constructs on the overexpression of MGR. We chose human embryonic kidney 293 (HEK293) cells for these experiments as they express MGR at low levels (van Laere *et al*, 2003), making it easier to observe effects due to overexpression. Consistent with low expression of MGR in HEK293 cells, we observed a smaller difference in luciferase activity between the q variant and Q variant construct as compared with that in C2C12 cells (Fig 3B). Co-transfection of a green fluorescent protein (GFP)-tagged MGR construct resulted in further differential repression of luciferase activity, with significantly stronger repression of the q variant construct as compared with that of the
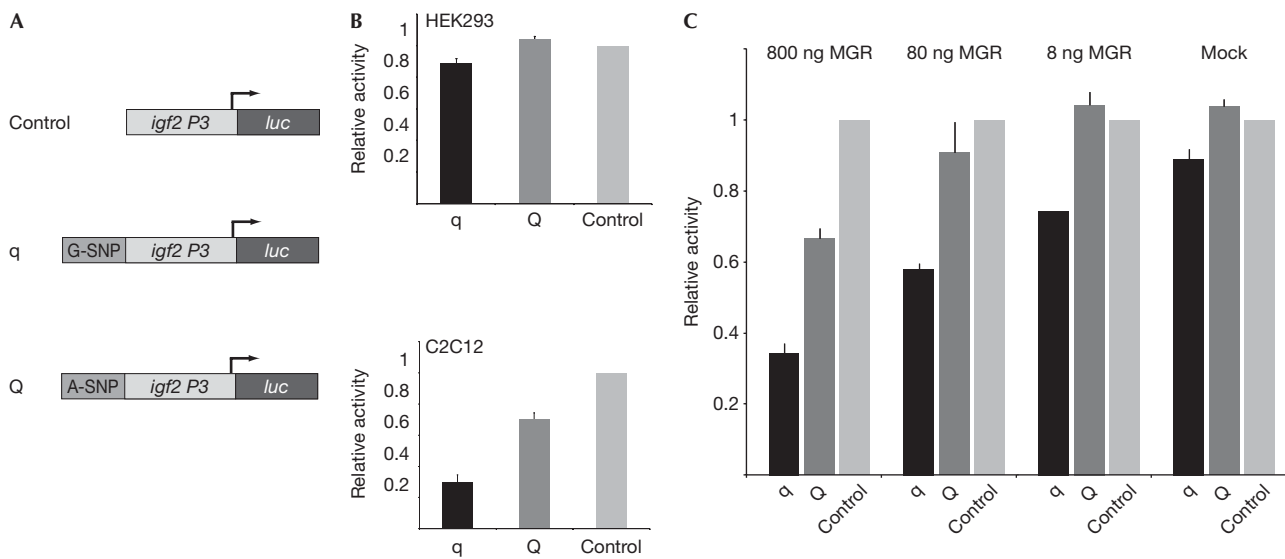
Fig 3 | Muscle growth regulator functions as a repressor for Igf2. (**A**) The schema of the transactivation constructs. (**B**) Transfections of reporter gene constructs in HEK293 cells show only slight repression for the q allele, whereas in C2C12 myoblast cells the reporter construct shows stronger repression at the q allele together with partial repression of the Q allele. (**C**) Co-transfection of MGR results in dosage-dependent repression of the q allele (for 80 ng GFP–MGR repression at the q variant is $30 \pm 3\%$ compared with that at the Q variant, which is $9 \pm 4\%$), which can mimic the repression pattern observed in C2C12 cells at 800 ng co-transfected MGR. GFP, green fluorescent protein; HEK, human embryonic kidney; MGR, muscle growth regulator.

Q variant construct in a dosage-dependent manner ($P < 0.0011$). When a sufficient amount of repressor was present, reporter gene activities were comparable to those in C2C12 cells, which express the repressor at higher endogenous levels (Fig 3C).

MGR binds to Igf2 locus and regulates Igf2 levels

We performed bisulphite sequencing of the Igf2 region in C2C12 to validate that the CpG pairs adjacent to the palindromic sequence, which we have shown to abrogate MGR binding, are not methylated completely. We observed only moderate levels of methylation in C2C12 (in contrast to NIH3T3 and D32) cells—which is consistent with a role in muscle development given that methylation of the Igf2 locus inhibits MGR binding (Fig 4A).

We investigated the effect on Igf2 mRNA levels during MGR knockdown by using three independent endoribonuclease-prepared small interfering RNA (esiRNAs; Kittler *et al*, 2007). Consistent with its role as a repressor of Igf2 expression, we observed threefold upregulation of endogenous Igf2 mRNA levels in proliferating cells on MGR knockdown (Fig 4B).

Attempts to stably overexpress MGR in C2C12 myoblasts proved unsuccessful, probably indicating that high cellular levels of this protein are not tolerated. We therefore used the bacterial artificial chromosome (BAC)-based TransgeneOmics approach (Kittler *et al*, 2005; Poser *et al*, 2008) to generate clonal C2C12 myoblast cell lines with a stably integrated additional MGR gene. As MGR is still expressed from its endogenous promoter, maintaining all *cis*-regulatory elements, our C2C12 clones with localization and affinity purification (LAP)-tagged MGR allowed us to study the effects of mild MGR overexpression. Thirteen clonal lines expressing MGR-LAP were established, all of which showed growth retardation compared with wild-type C2C12 cells,

consistent with the expected effect of MGR on muscle growth (van Laere *et al*, 2003; Stinckens *et al*, 2007). The doubling time was increased markedly from 16 h for wild-type C2C12 cells to about 32 h for C2C12 BAC cells (Fig 4C). By contrast, several clonal murine embryonic stem cell lines stably expressing the same BAC construct did not show obvious growth differences compared with wild-type embryonic stem (ES) cells (Fig 4D). This seems to be due to lack of binding: whereas a 140 kDa fusion protein is observed in the stably integrated BAC in C2C12, in murine stem cells all clones express a C-terminal fragment of 110 kDa, which lacks the nuclear localization signal and DNA-binding domain (Fig 4E). We performed chromatin immuno-precipitation (ChIP) analysis with an MGR-LAP-tagged clonal C2C12 cell line to determine whether MGR binds to the Igf2 locus *in vivo*. By using two GFP-affinity resins (nanotrap and antibody) we detected the enrichment of MGR-GFP at the SNP-containing region of the Igf2 locus (Fig 4F).

DISCUSSION

Our quantitative, high-accuracy proteomics screen identified unambiguously a single candidate as binding differentially to the quantitative trait locus. Independent lines of evidence further support the functional identity of the protein: MGR binds differentially to the q variant of a SNP and functions as a repressor; mild overexpression induces the severe growth pheno-type in C2C12 myoblast cells and knockdown of MGR results in upregulation of Igf2. Thus, we conclude that our unbiased quantitative proteomics approach has identified MGR, a domes-ticated, transposable element of the hAT family that binds to the *IGF2* G3072A SNP and is responsible for increased muscle tissue in many commercially bred pigs. In wild-type pigs, the *IGF2* locus
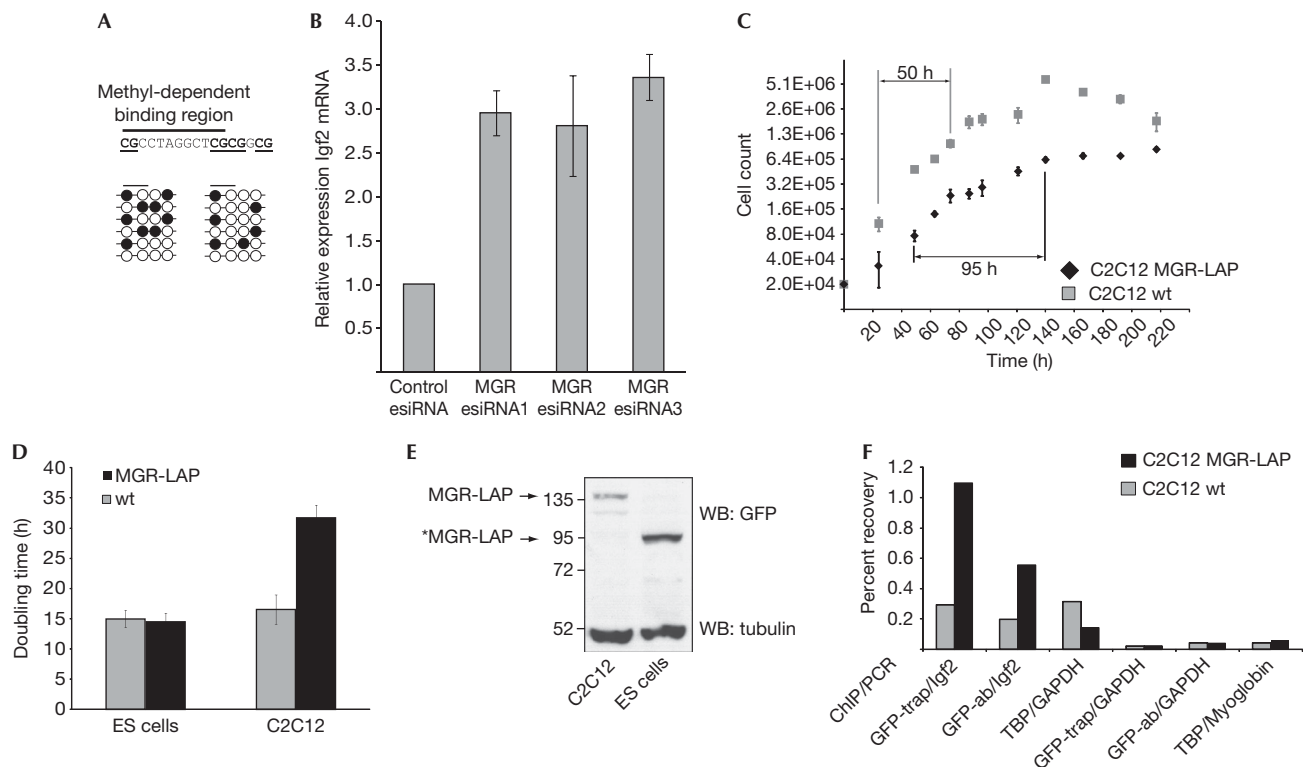
Fig 4 | Functional validation of muscle growth regulator *in vivo*. (A) Bisulphite sequencing of the binding region of MGR in C2C12 cells. (B) RNA interference with three independent esiRNAs shows threefold Igf2 upregulation in proliferating, subconfluent C2C12 cells. (C,D) Growth curves of C2C12-LAP and wild-type cell lines show an at least twofold increase in doubling times for the MGR-LAP cell line. An embryonic stem cell line containing the identical BAC construct has no significant growth differences compared with the wild type. (E) Immunoblot of C2C12-LAP and mES-LAP shows expression of an amino-terminal truncated version of MGR-LAP (indicated by the asterisk) missing the nuclear localization signal and the DNA-binding domain. (F) ChIP shows tenfold (GFP-nanotrap) and sixfold (antibody) higher recovery in C2C12-LAP in comparison with that in wild-type cells (after normalization on TBP/GAPDH). BAC, bacterial artificial chromosome; ChIP, chromatin immunoprecipitation; ES cells, embryonic stem cells; esiRNA, endoribonuclease-prepared small interfering RNA; GAPDH, glyceraldehyde 3-phosphate dehydrogenase; GFP, green fluorescent protein; MGR, muscle growth regulator; TBP, TATA box binding protein; WB, western blot; wt, wild type.

is bound by the repressor, whereas in pigs with the G3072A allele, the repressor cannot bind, leading to enhanced Igf2 expression and muscle mass. Igf2 overexpression in a transgenic mouse leads to prenatal overgrowth (Sun *et al*, 1997); however, the action of MGR might be restricted to muscle cells because of the expression of functional MGR and low methylation at this locus as compared with, for example, liver cells (van Laere *et al*, 2003). Beyond its interest for explaining the functional mechanism for the quantitative trait locus associated with lean pigs, the high conservation of MGR suggests that it might have important roles in other animals and possibly in human diseases characterized by muscle hypotrophy. Furthermore, our study illustrates that unbiased quantitative proteomics is an effective method by which to extend the results of genetic screens to the functional biochemical level.

METHODS

DNA pulldown. A 25 μg weight of annealed, concatenated and biotinylated DNA probes (forward 5′-TTCAAAACTGCAGGATC CTTCGCCTAGGCTC[A/G]CAGCGCGGGAGCGA-3′ and reverse

5′-AATCGCTCCCGCGCTG[T/C]GAGCCTAGGCGAAGGATCCTG CAGTTTTG-3′, *Pst*I restriction site) was bound to 75 μl of Dynabeads MyOne C1 (Invitrogen, Karlsruhe, Germany). Excess oligonucleotides were removed and beads were incubated with 400 μg of SILAC-labelled nuclear extracts (supplementary information online) in protein binding buffer (150 mM NaCl, 50 mM Tris–HCl (pH 8), 0.5% NP-40, 10 mM $MgCl_2$, protease inhibitor cocktail; Roche, Basel, Switzerland). After 2 h of slight agitation at 4 °C, the beads were washed three times, combined and boiled, and eluted proteins were separated by one-dimensional denaturing SDS-polyacrylamide gel electrophoresis (SDS-PAGE; Novex 4–12% gradient gel; Invitrogen). Mass spectrometric sample preparation was performed as described previously (Vermeulen *et al*, 2007).

For western blot analysis, non-labelled nuclear HEK293 lysates containing the recombinantly expressed protein were used similarly, except after SDS-PAGE when proteins were transferred to a Protran 85 membrane (Whatman, Maidstone, UK) and probed with a monoclonal GFP antibody (monoclonal mouse; Roche) in combination with a horseradish peroxidase-coupled mouse

secondary antibody (GE Healthcare, Munich, Germany). Proteins were detected by chemo-luminescence by using the ECL Detection kit (GE Healthcare). Pulldowns were conducted with the following probes: porcine (see above); murine (5′-TTCAAAACTGCA GGATCCTTCGCCTAGGCTC[A/G]CGGCGTCTGAGCGA-3′ and 5′-AATCGCTCAGACGCCG[T/C]GAGCCTAGGCGAAGGATCCTG CAGTTTTG-3′) and human (5′-TTCAAAACTGCAGGATCCTTCG CCTAGGCTC[A/G]CAGCGCGGAGGCGA-3′ and 5′-AATCGCCT CCGCGCTG[T/C]GAGCCTAGGCGAAGGATCCTGCAGTTTTG-3′).

Electrophoretic mobility-shift assays. Electrophoretic mobility-shift assay was performed as described previously (van Laere et al, 2003) with 1 pmol of Cy3-labelled oligonucleotide and E. coli lysate containing the recombinant murine protein. DNA oligos for porcine Igf2 (5′-TTCAAAACTGCAGGATCCTTCGCCTAGGCTC[A/G] CAGCGCGGGGAGCGA-3′ and 5′-(Cy3)TCGCTCCCGCGCTG[T/ C]GAGCCTAGGCGAAGGATCCTGCAGTTTTGAA-3′) and murine Igf2 (5′-GATCCTTCGCCTAGGCTC[A/G]CGGCGTCTGAGC GA-3′ and 5′-(Cy3)TCGCTCAGACGCCG[T/C]GAGCCTAGGCGA AGGATC-3′) were obtained with 5′-Cy3 modification from Metabion (Martinsried, Germany). DNA–protein complexes were separated on a 6% native polyacrylamide gel in $0.5\times$ TBE using 20 W for 2 h. The gel was screened by using Cy3 excitation for fluorescence detection and an FLA-7000 scanner (Fujifilm, Düsseldorf, Germany). Images were scanned into the supplied Multi Gauge Software and exposure was adjusted manually.

Transactivation assay. We constructed the luciferase-reporter vectors containing either the A variant or the G variant by cloning the murine regulatory region (55 bp upstream and 38 bp downstream from the SNP) upstream from the murine Igf2 promoter (from -227 to $+40$ relative to transcription start site; Sloning, Puchheim, Germany) and subcloning into pGL3-basic (Promega, Madison, WI, USA). C2C12 or HEK293 cells were grown to 80% confluence and transfected with 0.8 µg of the luciferase reporter and 80 ng of Renilla control vector (hRL-TK; Promega) by using Lipofectamine 2000 (Invitrogen). For co-transfection experiments variable amounts of GFP-MGR were transfected. Cells were incubated for 24 h before lysis and luciferase activity was detected by using a dual-luciferase Reporter Assay System (Promega) measured at a Mithras LB 940. The results are based on three independent experiments performed in triplicate; for statistical analysis standard deviations and t-test values were determined.

Monoclonal LAP cell line. The BAC RP23-318E14 harbouring MGR was obtained from the BACPAC Resource Center (http:// bacpac.chori.org). A LAP cassette (Cheeseman & Desai, 2005) was inserted as a C-terminal fusion by recombineering (Zhang et al, 2000; Gene Bridges, Dresden, Germany). The isolated BAC DNA was transfected and selected for stable integration as described (Poser et al, 2008) in C2C12 and mouse embryonic stem cells. Positive transfected cells were enriched by fluorescence-activated cell sorting and individual clones were selected. To assess doubling times, equal numbers of a clonal line were seeded in 24-well plates and three replicates were measured at different time points (COUNTESS; Invitrogen).

esiRNA transfection. For all transfections, 100,000 C2C12 cells were seeded per 3.5 cm² dish and incubated overnight 1 day before transfection. For esiRNA transfection the next day, 10 µl of Lipofectamine RNAiMax (Invitrogen) was diluted in 250 µl OptiMEM (Invitrogen) and incubated for 5 min at room temperature (25 °C). In a separate tube, 2 µg of esiRNA was diluted in

250 µl of OptiMEM. The solutions were combined, mixed and incubated for 20 min at room temperature, after which the transfection mix was spread evenly over the dish.

ChIP. C2C12 cells and C2C12 cells expressing MGR-GFP were crosslinked for 20 min at room temperature using 1% (v/v) formaldehyde. Chromatin was isolated and sonicated as described (Brinkman et al, 2007). Chromatin from about 1 million cells was subjected to IP using a polyclonal GFP binder resin (U Rothbauer) or a polyclonal rabbit antibody against GFP (Abcam AB0290), as described (Brinkman et al, 2007). As a control for chromatin preparations, chromatin from wild-type C2C12 and GFP-MGR cells was also subjected to IP using an antibody against TATA box binding protein (SL30). Following IP, eluted protein–DNA complexes were decrosslinked and purified DNA was subjected to quantitative PCR using the following primer sets: Igf2, forward CCTGGCCCCGACTTCGCCTA, reverse TCGGAGAGGGA GGCGGAACG; GAPDH, forward ACAGGGAGGAGCAGAGA GCA, reverse TCCTATAAATACGGACTGCAGCC; myoglobin second exon, forward AGAAGGGACAACATGCTGCC, reverse TTGTGCTTGGTGGCGTGT. ChIP efficiency was determined by plotting the percentage of input chromatin that was immunoprecipitated in the experiments.

Supplementary information is available at EMBO reports online (http://www.emboreports.org).

CONFLICT OF INTEREST
The authors declare that they have no conflict of interest.

REFERENCES
Aravind L (2000) The BED finger, a novel DNA-binding domain in chromatin-boundary-element-binding proteins and transposases. Trends Biochem Sci 25: 421–423
Brinkman AB, Pennings SW, Braliou GG, Rietveld LE, Stunnenberg HG (2007) DNA methylation immediately adjacent to active histone marking does not silence transcription. Nucleic Acids Res 35: 801–811
Cheeseman IM, Desai A (2005) A combined approach for the localization and tandem affinity purification of protein complexes from metazoans. Sci STKE 2005: pl1
Cooper GM, Stone EA, Asimenos G, Green ED, Batzoglou S, Sidow A (2005) Distribution and intensity of constraint in mammalian genomic sequence. Genome Res 15: 901–913
Estelle J, Mercade A, Noguera JL, Perez-Enciso M, Ovilo C, Sanchez A, Folch JM (2005) Effect of the porcine IGF2-intron3-G3072A substitution in an outbred Large White population and in an Iberian x Landrace cross. J Anim Sci 83: 2723–2728
Feschotte C, Pritham EJ (2007) DNA transposons and the evolution of eukaryotic genomes. Annu Rev Genet 41: 331–368
Jungerius BJ, van Laere AS, Te Pas MF, van Oost BA, Andersson L, Groenen MA (2004) The IGF2-intron3-G3072A substitution explains a major imprinted QTL effect on backfat thickness in a Meishan x European white pig intercross. Genet Res 84: 95–101

Kempken F, Windhofer F (2001) The hAT family: a versatile transposon group common to plants, fungi, animals, and man. *Chromosoma* **110:** 1–9

Kittler R, Pelletier L, Ma C, Poser I, Fischer S, Hyman AA, Buchholz F (2005) RNA interference rescue by bacterial artificial chromosome transgenesis in mammalian tissue culture cells. *Proc Natl Acad Sci USA* **102:** 2396–2401

Kittler R *et al* (2007) Genome-wide resources of endoribonuclease-prepared short interfering RNAs for specific loss-of-function studies. *Nat Methods* **4:** 337–344

Mittler G, Butter F, Mann M (2009) A SILAC-based DNA protein interaction screen that identifies candidate binding proteins to functional DNA elements. *Genome Res* **19:** 284–293

Oczkowicz M, Tyra M, Walinowicz K, Rozycki M, Rejduch B (2009) Known mutation (A3072G) in intron 3 of the IGF2 gene is associated with growth and carcass composition in Polish pig breeds. *J Appl Genet* **50:** 257–259

Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* **1:** 376–386

Poser I *et al* (2008) BAC TransgeneOmics: a high-throughput method for exploration of protein function in mammals. *Nat Methods* **5:** 409–415

Rubin E, Lithwick G, Levy AA (2001) Structure and evolution of the hAT transposon superfamily. *Genetics* **158:** 949–957

Stinckens A, Van den Maagdenberg K, Luyten T, Georges M, De Smet S, Buys N (2007) The RYR1g.1843C>T mutation is associated with the effect of the IGF2 intron3-g.3072G>A mutation on muscle hypertrophy. *Anim Genet* **38:** 67–71

Sun FL, Dean WL, Kelsey G, Allen ND, Reik W (1997) Transactivation of Igf2 in a mouse model of Beckwith–Wiedemann syndrome. *Nature* **389:** 809–815

van Laere AS *et al* (2003) A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. *Nature* **425:** 832–836

Vermeulen M, Mulder KW, Denissov S, Pijnappel WW, van Schaik FM, Varier RA, Baltissen MP, Stunnenberg HG, Mann M, Timmers HT (2007) Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell* **131:** 58–69

Vermeulen M, Hubner NC, Mann M (2008) High confidence determination of specific protein–protein interactions using quantitative mass spectrometry. *Curr Opin Biotechnol* **19:** 331–337

Zhang Y, Muyrers JP, Testa G, Stewart AF (2000) DNA cloning by homologous recombination in *Escherichia coli*. *Nat Biotechnol* **18:** 1314–1317

Zhao X, Xuan Z, Zhang MQ (2007) Boosting with stumps for predicting transcription start sites. *Genome Biol* **8:** R17