Supplementary Results

A) Lever screen of 4 known myogenic regulatory motifs across 101 myogenic gene sets

As an initial positive control analysis, we applied Lever to systematically analyze the 101 myogenic gene expression clusters and GO categories when considering the four known myogenic motifs MRF, MEF2, SRF and Tead. We found that 41 out of the 101 gene sets showed significant enrichment ($Q \leq 0.05$) for at least one Boolean combination of these four motifs (**Supplementary Figure 6** online; **Supplementary Table 3b** online). Nearly all gene sets that showed statistically significant enrichment ($Q \leq 0.05$) for combinations of these four motifs were composed of up-regulated genes, consistent with the known functions of the corresponding TFs as transcriptional activators.

B) Experimental validation of CRMs predicted by PhylCRM

We first verified by Q-RT-PCR that these seven genes were up-regulated during differentiation (**Supplementary Figure 8** online). Western blot analyses confirmed that these myogenic TFs were differentially expressed at the protein level during differentiation (**Supplementary Figure 9** online). Next, chromatin immunoprecipitation (ChIP) assays followed by region-specific quantitative PCR (see **Methods**) showed that four of the six candidate CRMs were significantly enriched for binding by MEF2 ($P \leq 0.05$), MyoD ($P \leq 0.05$) and myogenin ($P \leq 0.005$) (**Figure 4b**). Positive control CRMs were also significantly enriched for binding by these TFs, while negative control regions were not (**Figure 4b**). Two of these four bound regions were also significantly occupied by SRF ($P \leq 0.05$) during differentiation. Interestingly, of the six tested CRMs, the four that showed significant binding by MEF2, MyoD, and myogenin were the ones that are located next to genes involved in sarcomeric function, whereas the two that did not show significant binding by these factors are not. Although this does not tell us what sequence features distinguish the active from the inactive CRMs, it does suggest that the choice of the likely target gene sets is important in predicting CRMs that are active in a given

condition (here, myogenic differentiation).

We performed luciferase assays for the four novel, candidate CRMs that were enriched for *in vivo* TF binding. All four of these candidate CRMs resulted in statistically significant ($P \leq 0.05$) activation of luciferase expression during myogenic differentiation (**Figure 4c**). In contrast, these same CRMs did not result in increased luciferase activity in either fibroblasts or lens epithelial cells (**Figure 4c**). To further validate that these four candidate CRMs drive expression specifically in response to myogenic differentiation, we disrupted myogenic differentiation by shRNA knockdown of MEF2D (one of two MEF2 isoforms up-regulated in myotubes), myogenin (the most up-regulated MRF member), or SRF (**Supplementary Figure 10** online). Knockdown of myogenin significantly reduced ($P \leq 0.05$) transcriptional activity of all four predicted human CRMs positive for luciferase reporter activity in C2C12 myotubes (**Supplementary Figure 11a** online), while knockdown of SRF or MEF2D reduced the transcriptional activity of different subsets of these CRMs (**Supplementary Figure 11b,c** online). We note as a caveat that this reduced luciferase activity could potentially have been due to indirect effects involving some other TF under the control of the myogenic regulators knocked down by the shRNAs. In each case the level of luciferase activity was proportional to the amount of TF knockdown for a given shRNA clone (**Supplementary Figures 10 and 11** online).

Finally, we tested the sufficiency of the MRF AND MEF2 motif combination for CRM activity by generating a synthetic CRM containing consensus MRF and MEF2 binding sites arranged as in our newly discovered *ACTA1* CRM, but in the context of the *MGLL* negative control flanking sequence (see **Methods**). This synthetic CRM failed to drive expression in a luciferase reporter construct, suggesting that there are further sequence requirements aside from the MRF and MEF2 motifs (**Supplementary Figure 12** online). We anticipate that further computational analyses with more candidate regulatory motifs, combined with further experimental testing, will help to identify additional sequence features that may be important for CRM activity.