The instructions for generating and sustaining all life forms are encoded within each life form’s genome. A major challenge in the postgenomic sequencing era has been assigning functions to the hundreds of thousands of noncoding sequence elements within our nuclear genome that are homologous to sequences in other species. These snippets of DNA have been hypothesized to impart information for initiating DNA replication, for inter- and intrachromosomal recombination, for structural integration of the genome into the surrounding nucleoskeleton, and for transcriptional regulation of gene expression. The latter regulatory elements have been of particular interest inasmuch as variations among our own species, and to some extent between species, are thought largely to be a function of differences in the timing, duration, and intensity of gene expression. Moreover, the explosive rise largely to be a function of differences in the timing, duration, and intensity of gene expression. Moreover, the explosive rise in single nucleotide polymorphism (SNP) association studies has generated a mounting number of noncoding SNPs whose functions, while poorly defined at this time, will undoubtedly include the regulation of gene expression. Finally, the continued discovery of regulatory elements controlling gene expression will augment our “genomic tool box” of reagents for expressing and inactivating genes in a context-dependent manner. Elucidating the function of all regulatory elements in our genome is therefore a critically important endeavor from both a clinical and basic science perspective.

The 3 muscle types display unique patterns of gene expression; however, during development and in some pathological states there is overlap in gene expression profiles suggesting a common mode of regulation (Figure 1). For example, the majority of cyto-contractile genes expressed in each muscle cell type are under direct control of the widely expressed transcription factor, serum response factor (SRF). SRF self-dimerizes and binds to a 10-bp sequence known as a CArG element or CArG box (Figure 2). CArG boxes are found in the 5′ promoter and intronic region of a rising number of cyto-contractile genes. Based on 20 years of DNA-protein and promoter analyses, as well as comparative genomics, we now recognize that SRF may potentially bind to 1216 permutations of a CArG box, with CCTTATATGG emerging as a consensus sequence (Figure 2). Recent genome-wide studies have further advanced our understanding of the base sequence character of CArG elements and have greatly expanded the so-called CArGome. As of this writing, more than 200 CArG boxes controlling expression of some 170 mammalian SRF target genes have been identified with more than 300 hypothetical CArG boxes awaiting wet-laboratory validation.

SRF possesses relatively weak transcriptional activity, but binds to any one of 56 cofactors that potently activate target gene expression, mainly through alterations in chromatin permissive for DNA transcription. Many SRF cofactors exhibit cell-restricted patterns of gene expression during development and postnatal life. One of the more powerful cell-restricted SRF cofactors is myocardin (Myocd), which was first cloned in a bioinformatic screen for cardiac-specific genes. Expression of Myocd is highly specific for cardiac and SMCs, with transient expression in developing skeletal muscle precursors. Thus, Myocd forms a ternary complex with SRF-bound CArG boxes and, through its association with a variety of other coregulators of gene expression, directs expression of cardiac and SMC cyto-contractile genes. Though cardiac genes are induced when Myocd is ectopically expressed in nonmuscle cells, little evidence of a structural or functional cardiac muscle phenotype is manifest. In contrast, Myocd orchestrates structural, biochemical, and physiological characteristics of SMCs. Thus, Myocd appears to be the SMC equivalent of MyoD, the original master regulator of the skeletal muscle phenotype.

SMCs are defined by a molecular signature of gene expression that includes genes encoding contractile, cytoskeletal, ion channel, transcription factor, and signaling proteins, all of which are essential to carry out the unique function of this cell type. The regulatory regions of many of these genes have been characterized both in vitro and in vivo and more than half contain functional CArG elements. Now, in this issue of Circulation Research, Petit et al report the discovery of an alternative form of the SMC-specific gene, LIM domain containing preferred translocation partner in lipoma (aka Lpp), that appears to be under direct control of CArG-SRF-Myocd ternary complexes. Lpp protein expression was shown previously to be highly specific for SMCs where, in association with vinculin at peripheral dense bodies, it mediates cell migration. Petit et al sought to define whether any functional CArG boxes reside in or around the 588-kb mouse Lpp locus using a bioinformatics approach based on CArG nucleotide frequencies (Figure 2). A total of 35 CArG elements were found over the interrogated sequence, a number surprisingly lower than the theoretical frequency of 1 every 910 base pairs of DNA sequence. Three of the 35 CArG elements were found to be homologous with corresponding CArGs in several other species, though their position is more than 50 kb away from the annotated start site of Lpp transcription. Because virtually all functional CArG elements reside within a 4-kb window of transcription start sites, Petit
et al searched for possible internal promoters in the Lpp gene. Their presupposition of an internal promoter is supported by recent knowledge demonstrating a much more complex transcriptome than we ever imagined. Using various genomic algorithms and RT-PCR, the authors discovered an additional promoter inside intron 2 of the mouse Lpp gene that directs expression of a SMC-specific transcript. Because the first 4 exons of mouse Lpp are noncoding, the alternate SMC-specific transcript encodes for an identical LPP protein as that derived from the more proximal promoter.

The intronic Lpp promoter resides \( \approx 50 \) kb downstream from the proximal promoter region where at least 5 nonhomologous CArG elements are found. None of the latter CArGs appears to bind SRF using ChIP though it would be important to extend these studies to additional promoter assays. Interestingly, the human LPP proximal promoter contains 4 CArG boxes over a 350-bp genomic interval, a density 10-fold greater than that predicted by chance. Whether these CArGs bind SRF and respond to SRF-Myocd as permissive for protein-DNA binding (eg, hypoacetylation) is unknown. In vitro reporter assays revealed intrinsic promoter activity for the intronic Lpp promoter that was partially dependent on CArG8. Both SRF and Myocd activated the intronic Lpp promoter in a CArG8-dependent manner suggesting, at least in vitro, that Lpp is a direct target of SRF-Myocd. These results are congruent with the original finding that Lpp mRNA is induced with Myocd. Petit et al further corroborated these findings with elegant expression studies using embryonic stem cells null for SRF as well as various tissues where SRF was deleted specifically in SMCs. In both cases, the intronic Lpp promoter-driven transcript is sharply attenuated, but could be rescued on ectopic SRF expression.

The study by Petit et al further expands the mammalian CArGome and the molecular signature of SMC lineages. Future work will require a thorough analysis of both the proximal and intronic Lpp promoters in transgenic mice. In addition, the presence of a microRNA (miR-28) within intron 8 of the Lpp gene should be of some interest given the number of SRF-responsive microRNAs. Finally, ongoing efforts continue to ascertain whether any sequence variants within or adjacent to the more than 500 CArGs identified in our genome are linked to altered target gene expression and human disease.

Sources of Funding
The author’s work on this subject is funded through National Institutes of Health grant HL-62572.

Disclosures
None.

References


**Key Words:** serum response factor / myocardin / promoter / knockout / bioinformatics
Deck of CArGs
Joseph M. Miano

doi: 10.1161/CIRCRESAHA.108.179978

_Circulation Research_ is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2008 American Heart Association, Inc. All rights reserved.
Print ISSN: 0009-7330. Online ISSN: 1524-4571

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circres.ahajournals.org/content/103/1/13

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in _Circulation Research_ can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to _Circulation Research_ is online at:
http://circres.ahajournals.org/subscriptions/