Regulation Of Transcription And Regulatory Networks For Muscle Growth

A. Reverter[*], N.J. Hudson[*], Q. Gu[*] and B.P. Dalrymple[*]

Introduction

The advent of microarray gene expression technology has provided animal scientists with an unprecedented ability to profile the transcriptional changes during skeletal muscle growth. With respect to meat quality, most of the effort has concentrated on the understanding of fat and energy metabolism (reviewed by Hausman *et al*. (2009)). Graugnard *et al*. (2009) explored the network among 31 genes associated with aspects of adipogenesis and energy metabolism in bovine skeletal muscle and in response to two distinct diets. Also, Freyssenet (2007) reviewed the roles that energy-sensing molecules and mitochondria have in the regulation of gene expression in muscle. However, other mechanisms such as cell cycle, glycolysis, extra-cellular matrix, ribosomal proteins and the immune system play a significant role in development, and this role can work in a tissue-specific manner.

Hudson *et al*. (2009a) reported various functional modules underpinning the transcriptional regulation of bovine skeletal muscle. The authors integrated a total of six gene co-expression networks, each developed using the PCIT algorithm (Reverter and Chan (2008)), and proposed a Module-to-Regulator heuristic by which those transcription factors (TF) with the highest average absolute correlation co-expression with the genes present in each module are deemed to be the relevant regulators. However, this Module-to-Regulator approach failed to capture some well-known regulators of muscle fibre type composition, and the use of more sophisticated methods such as the differential wiring approach of Hudson *et al*. (2009b) or the Regulatory Impact Factor (RIF) metrics of Reverter *et al*. (2010) was suggested. Importantly, while the RIF metrics were designed to be computed for a given experimental contrast (eg. Breed 1 vs. Breed 2), there is no reason why the concepts of differential wiring (case of RIF1) and change in predictive ability (case of RIF2) cannot be applied across genes in a module, irrespective of their differential expression in a contrast of interest.

In this study, we summarize the recent developments in gene co-expression networks for skeletal muscle and present a curated list of 190 module genes across six functionally coherent modules that are of relevance to muscle growth and development. We then undertake RIF analyses on over one thousand regulators against the module genes, and apply PCIT on the regulators' RIF metrics across the six modules to generate a network of regulators. Finally, we discuss our approach and findings in the context of the known biology that's being recapitulated, the novel hypotheses that are being generated, and the potential for a myriad of rational modifications to our approach that could provide additional insights.

[*] Computational and Systems Biology, CSIRO Livestock Industries, St. Lucia, Qld 4067, Australia

Material and methods

Co-expression networks for bovine skeletal muscle. In Hudson *et al*. (2009a) we inferred the transcriptional landscape of bovine skeletal muscle by integrating co-expression networks. This work is central because it represents the culmination of previous efforts by the same group in the area of construction of gene interaction and regulatory networks in Bovine skeletal muscle (Reverter *et al*. 2005, 2006). In brief, the authors used the expression profile of 13,094 bovine genes across 26 experimental conditions to reverse engineer six interlaced networks: Pre-natal, Post-natal, Piedmontese, Wagyu, Diets, and Overall. The edges of the network (represented by gene to gene co-expression correlations) that were significant across the six networks were used to build the so-called 'Always Correlated' landscape. Upon careful exploration of the Always Correlated landscape using the visualization suite of tools of Cytoscape (Shannon *et al*. (2003); www.cytoscape.org), a number of functionally coherent modules were identified.

Module genes. For the present study, the functional modules reported by Hudson *et al*. (2009a) were subjected to further scrutiny to generate a curated list of module genes. This additional examination was based on hierarchical cluster analysis using PermutMatrix software (Caraux and Pinloche, 2005) as well as on the molecular function gene ontology term (www.geneontology.org). We focused our attention on six modules as follows: Cell Cycle, Fat Metabolism, Immune System, Mitochondria, Muscle/Glycolysis, and Ribosome.

Regulatory Impact Factor (RIF) analyses and the network of regulators. The RIF algorithm was initially introduced to identify the critical transcription factors (TF) from expression data (Hudson *et al*. 2009b). That work was based on the specific case of myostatin (MSTN), which demonstrates a previously identified, trait-associated genetic variant in one breed of cattle (Piedmontese), but not another (Wagyu). The RIF identified MSTN as the regulatory factor with the highest evidence of contributing to differential expression, across a developmental time-course experiment, in the absence of any demonstrably differential expression of the regulator itself. In Reverter *et al*. (2010), we tested whether this approach was generally applicable, testing four, appropriately diverse, experimental datasets. The strength of RIF lies in its ability to simultaneously integrate three sources of information into a single measure: (1) The differential wiring (DW) defined as the change in correlation between the TF and the differentially expressed (DE) genes; (2) The amount of differential expression of the DE genes; and (3) The abundance of DE genes. As a result, RIF assigns an extreme score to those TF that are consistently most differentially co-expressed with the highly abundant and highly DE genes (RIF1), and to those TF with the most altered ability to predict the abundance of DE genes (RIF2). However, while RIF was conceived to contrast TF against the DE genes, its formulae are equally applicable to contrasting TF against genes in a given module. The intent here is to identify the key regulators of a given functionally coherent module, irrespective of the differential expression of the genes that comprise the module. To this respect, two alternative measures of RIF are computed as follows (Reverter *et al*. 2010):

$$\mathrm{RIF1}_i = \frac{1}{\mathrm{n_{de}}} \sum_{j=1}^{j=\mathrm{n_{de}}} \hat{a}_j \times \hat{d}_j \times \mathrm{DW}_{ij}^2 = \frac{1}{\mathrm{n_{de}}} \sum_{j=1}^{j=\mathrm{n_{de}}} \mathrm{PIF}_j \times \mathrm{DW}_{ij}^2 \qquad (1)$$

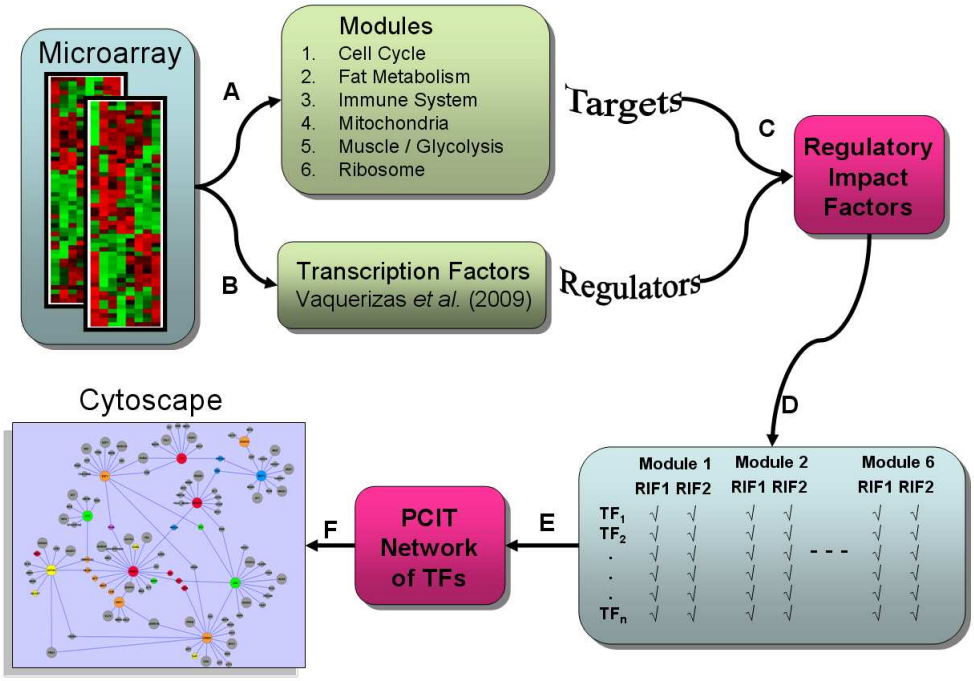

Figure 1: Schematic representation of the process used to generate a network for transcriptions factors (TF): Microarray data is surveyed to (A) identify modules of functionally related genes and (B) retrieve TFs. A Regulatory Impact Factor analysis (C) is undertaken contrasting the TFs as regulators with the module genes as targets. As result, a matrix is built with as many rows as TFs and 12 columns corresponding to RIF1 and RIF2 scores across the six modules (D). This matrix becomes the input to the PCIT algorithm (E) to reverse-engineer a network among TFs. Finally, the resulting network is visualized with Cytoscape (F) using attributes corresponding to the significance of the association of each TF to the functionally coherent modules

and $$\mathrm{RIF2}_i = \frac{1}{\mathrm{n}_{\mathrm{de}}} \sum_{j=1}^{j=\mathrm{n}_{\mathrm{de}}} \left[\left(e1_j \times r1_{ij}\right)^2 - \left(e2_j \times r2_{ij}\right)^2 \right], \quad (2)$$

where n_{de} is the number of DE genes (genes in a module in our context); $\hat{a}_j$ is the estimated average expression of the *j*-th module gene; $\hat{d}_j$ is the estimated differential expression (in the Piedmontese vs. Wagyu contrast) of the *j*-th gene; and DW is the differential wiring between the *i*-th TF and the *j*-th gene, and computed from the difference between $r1_{ij}$ and $r2_{ij}$, the co-expression correlation between the *i*-th TF and the *j*-th gene in Piedmontese and Wagyu samples, respectively.

A network of TFs. We interrogated the manually curated census of 1,391 TF supplied by Vaquerizas *et al*. (2009) and found that 1,017 of them were represented in our bovine

Table 1: Module Genes – The six functional modules profiled in this study

Module	N	Gene Symbols
1. Cell Cycle	49	*ANLN CCDC99 CCNA2 CCNB1 CCNB2 CCNF CDC2 CDC20 CDC25A CDC25B CDC6 CDCA2 CDCA3 CDCA5 CDCA8 CDKN3 CENPE CENPF CEP55 CHEK1 CKAP2 DLGAP5 E2F1 E2F2 E2F8 FOXM1 KCNQ1 KIAA0101 KIF15 KIF20A KIF23 KIF4A KIFC1 MND1 NDC80 PAPSS1 PBK PLK4 PRR11 PTTG1 RASSF4 RRM2 SMC2 SPC24 SPC25 TACC3 TOP2A TROAP UHRF1*
2. Fat	20	*ACSM1 ACSS2 ADIG ADIPOQ ADIPOR1 ADIPOR2 AGPAT2 CIDEA CIDEC DGAT2 FABP4 FADS1 FADS2 FBP1 LDLR PCK1 PIP5K1B PLIN PLS1 TUSC5*
3. Immune	22	*ASPA BOLA-DMB BOLA-DRA BSG C1R CD22 CD244 CD3E CD3G CD74 CTSW FCGR3A HLA-DRA IKZF3 IL2RB IRF1 IRF9 ITGAL LCP1 MYO1G PSMB8 PSMB9*
4. Mitochondria	33	*ACO2 APOO BRP44 COQ9 COX5A COX7A1 COX7B CYCS DLAT DLD ECHDC3 ECSIT ENDOG ES1 GBAS GOT2 MDH1 MDH2 MRPS36 NDUFA3 NDUFA5 NDUFB5 NDUFS2 NDUFS3 NDUFS7 NDUFV1 NDUFV2 NNT PDHA1 PDHX SLC25A3 SUCLA2 UQCRC1*
5. Muslce / Glycolysis	44	*ACTN3 AK1 AKR1B10 ATP2A1 BGLAP BIN1 CDH22 CIDEB CKM CMTM4 DHDH DHRS7C ENO3 GPI GRM6 HESX1 HRC IDH3A ITGB4 KCNG2 LAMB3 LZTR1 MACROD1 MTX3 MYBPC2 MYH1 MYLK2 MYOM2 MYOZ1 NEB NT5C2 PAX2 PGAM2 PGM1 PKM2 PYGM SLC16A3 SUV420H2 TBX15 TNNT3 TPI1 TPM1 TPM2 UBXN1*
6. Ribosome	22	*RPL11 RPL13 RPL13A RPL18 RPL18A RPL19 RPL23 RPL24 RPL35A RPL38 RPL5 RPL8 RPLP0 RPS11 RPS14 RPS15 RPS19 RPS24 RPS3 RPS5 RPS7 RPS9*

skeletal muscle gene expression dataset. The RIF analyses generated RIF1 and RIF2 scores for each TF and across the six modules. This (1,017 × 12) matrix of RIF scores was used as the input of the PCIT algorithm (Reverter and Chan (2008)) to obtain a network of TFs.

Overall schema. Figure 1 illustrates the schema of the work presented here. The gene expression microarray data of Hudson *et al*. (2009a) on bovine skeletal muscle is surveyed to identify module genes across a range of six functionally coherent modules. Simultaneously, the TF included in the microarray data are flagged based on the census described by Vaquerizas *et al*. (2009). A RIF analysis is performed by which TF and module genes are

deemed as regulators and targets, respectively. As a result, a matrix with as many rows as TF and 12 columns (from 2 RIF scores × 6 functional modules) is built and used as the input in the PCIT algorithm (Reverter & Chan 2008) to obtain a TF-TF network. Finally, the attributes of the resulting network are visualized using Cytoscape (Shannon *et al*. (2003)).

Results and discussion

Table 1 presents the composition of the functionally coherent modules explored in this study and comprising a total of 190 genes. Module size ranged from 20 to 49 genes for the Fat and the Cell Cycle modules, respectively. Reasons for varying module sizes can be attributed to the numerical difficulty, in terms of likelihood, to identify large cohesive modules as well as to the biological reasoning for larger modules to correspond to essential pathways (ie. embryonically lethal if seriously perturbed).

The PCIT network among 1,017 TFs contained 53,848 edges implying a cohesive structure with a 10.42 % clustering coefficient (ie. proportion of all possible connections found to be significant). The connectivity distribution showed a power law consistent with the scale-free property of most nodes having few connections and a few nodes having a large number of connections. The 10 (or ~1% of all TFs) most connected TFs (*OLIG1, POU2F1, SNW1, KLF7, FOXK1, MYF5, GABPAP, NFATC4, CRX* and *ZNF263*) had ≥ 197 connections each and were connected to 432 TFs (or ~43%). Of these, we highlight MYF5 as a well-known master regulator of myogenesis. The development and differentiation of skeletal muscle in mammals are controlled by a transcriptional cascade in which four myogenic regulatory factors play key roles (Buckingham 1992): *MYF5, MYOG, MYF6* (or *MRF4*) and *MYOD*.

Table 2: Top 5 transcription factors according to RIF1 or RIF2 in each module

Module	According to...	Transcription Factors
1. Cell Cycle	RIF1	*GATA3 HOXB5 TBL1X ZNF165 ECD*
	RIF2	*ID1 SMAD7 ARID3B GATA3 HES5*
2. Fat	RIF1	*MORF4L2 ZNF180 TBL1X NFIA ZDHHC5*
	RIF2	*GLIS1 HES5 ARID4A ESRRB* ***ATF7***
3. Immune	RIF1	*ZNF341 ZNF180 MGA MORF4L2* ***MSL2L1***
	RIF2	***TCEB3*** *ZFX* ***HOXD3*** *VDR* ***ZNF205***
4. Mitochondria	RIF1	*HOXB5 TBL1X GATA3 ECD GLIS1*
	RIF2	*GLIS1 HES5 GATA3 SMAD7 ID1*
5. Muscle/Glycolysis	RIF1	*TBL1X HOXB5 GATA3 HES5 MSX1*
	RIF2	*HES5 STAT4 SMAD7 ARID3B VDR*
6. Ribosome	RIF1	*SMAD7 ECD MSX1 ID1 ZNF165*
	RIF2	*MSX1 ARID3B ZNF165 IRF4 MORF4L2*

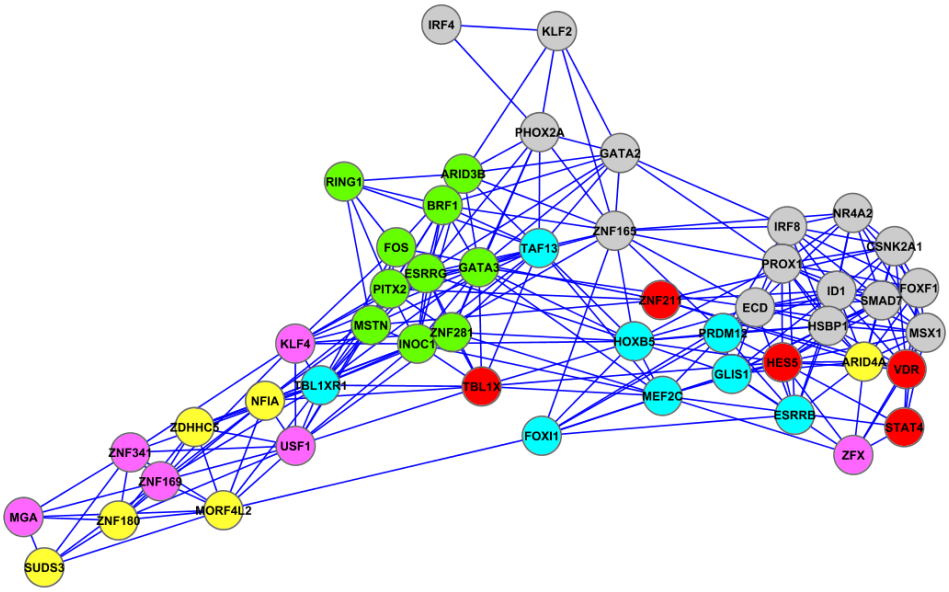

Figure 2: Network spanned by the top 50 transcription factors and across the six functionally coherent modules: Cell cycle (green), Fat (yellow), Immune (purple), Mitochondria (blue), Muscle/Glycolysis (red) and Ribosome (grey)

Table 2 lists the top 10 TFs (5 from each RIF metric) for each module. Because neither the modules nor the RIF metrics are completely independent from each other, this strategy yielded 28 unique TFs out of a total of 60 possible. Hence, many TFs were captured as 'top' in various modules and even within a module across the two RIF metrics (eg. case of *GATA3, SMAD7, MSX1*).

When TFs were ranked on the basis of their absolute RIF1 and RIF2 scores summed across all six modules and the top 50 (or ~5% of all TFs) selected, this new strategy captured 23 of the 28 TF reported previously, and with the 5 remaining indicated by bold face in Table2 (note four of them from the Immune System module).

We explored the network spanned by these 50 TF containing 272 edges (Figure 2). These highly ranked TF were allocated to one of the six modules according to the module in which they yielded the most extreme RIF score. With this scheme, we assigned 10, 6, 6, 8, 5, and 15 TFs to Cell Cycle, Fat, Immune, Mitochondria, Muscle/Glycolysis and Ribosome, respectively. The observed allocation was not statistically different from a random allocation of 50 items to 6 boxes (P = 0.5999; Chi-square with 5 d.f.). However, when this allocation was used as attributes in the visualization schema and the network displayed using the Spring Embedded layout, we obtained a landscape that preserved some of the original features of the module to regulator assignment. In this respect, TFs allocated to Ribosome, Cell Cycle and (to some extent) Mitochondria clustered in their own group, while TFs allocated to Fat and

Immune clustered together and somewhat separate from the other 4 modules. Interestingly, the five TFs allocated to Muscle/Glycolysis (*HES5*, *STAT4*, *TBL1X*, *VDR*, and *ZNF211*) were positioned in a rather scattered fashion. From this group, we highlight *VDR* (Vitamin D receptor). Working with mice, Endo *et al*. (2003) reported abnormal muscle development after deletion of *VDR*.

Equally suggestive is the allocation of *MEF2C* (myocyte-enhancer factor 2), *MSTN* (Myostatin) and *KLF4* (Kruppel-lilke factor 4) into the Mitochondria, Cell Cycle and Immune modules, respectively. Van Oort *et al*. (2006) showed that *MEF2* activity activates subsets of genes localized primarily to or functioning at the level of mitochondria. Thomas *et al.* (2000) showed that the hypertrophy of myotubes induced by Myostatin is a function of cell cycle. Finally, the immune function of *KLF4* as a suppressor of B and T cell proliferation has recently been established (Yusuf *et al.* (2008); Yamada *et al*. (2009)).

Finally, the finding that the top TFs allocated to the Fat module (yellow nodes in Figure 2) cluster together with those allocated to the Immune module (purple nodes in Figure 2) is fascinating given the growing body of evidence showing fat tissue as a mediator of immune system dysregulation (see for instance Pataky *et al*. (2009) and references therein).

Conclusion

While skeletal muscle can be simply described as a tissue composed of thousands multinucleated myofibers, efficient muscle growth and development is dependent on the coordinated action of multiple biological processes. The advent of microarray gene expression technology has allowed the confirmation of the traditional classification of muscles into 'red' and 'white', composed of slow and fast contractile fibers, respectively (Campbell *et al*. (2001)). Nevertheless, muscle remains a complex tissue to study using genomic approaches, because the molecular basis governing the biochemical and functional properties of muscle remain largely unknown.

The present study has shown that bringing together the forces of three approaches to the analysis of gene expression data (namely, differential expression, network inference, and differential wiring) allows us to correctly infer a network of regulators of cell cycle activity, glycolysis, mitochondrial transcription, adipogenesis and immune function, all within a skeletal muscle context.

Acknowledgements

We are grateful to Linda Café and Paul Greenwood who initiated and undertook the animal work that gave rise to the RNA samples that generated the datasets used in this study.

References

Buckingham (1992) *Trends Genet*. 8:144–149.

Campbell, W.G., Groden, S.E., Carlson, C.J., Pattison, J.S., Hamilton, M.T., and Booth, F.W. (2001) *Am. J. Physiol. Cell Physiol*. 280:C763–C768.

Caraux, G., and Pinloche, S. (2005). *Bioinformatics*, 21:1280–1281.

Endo, I., Inoue, D., Mitsui, T., Umaki, Y., Akaike, M., Yoshizawa, T., Kato, S., and Matsumoto, T. (2003) *Endocrinology*, 144:5135–5137.

Freyssenet, D. (2007) *J. Appl. Physiol*., 102:529–540.

Graugnard, D.E., Piantoni, P., Bionaz, M. *et al.* (2009) *BMC Genomics*, 10:142.

Hausman, G.J., Dodson, M.V., Ajuwon, K. *et al.* (2009) *J. Anim. Sci.*, 87:1218–1246.

Hudson, N., Reverter, A., Wang, Y.H. *et al.* (2009a) *PLoS ONE*, 4:e7249.

Hudson, N., Reverter, A., and Dalrymple B. (2009b) *PLoS Comput. Biol.*, 5:e1000382.

Pataky, Z., Bobbioni-Harsch, E., and Golay, A. (2009) *Exp. Clin. Endocrinol. Diabetes*, [Epub ahead of print].

Reverter, A., Barris, W., Moreno-Sánchez, N. *et al.* (2005) *Aust. J. Exp. Agric.*, 45, 821–829.

Reverter, A., Hudson, N.J., Wang, Y.H. *et al.* (2006) *Phys. Genomics*, 28, 76–83.

Reverter, A., and Chan, E.K.F. (2008) *Bioinformatics*, 24:2491–2497.

Reverter, A., Hudson, N., Nagaraj, S.H. *et al.* (2010) *Bioinformatics*, 26:896–904.

Shannon P, Markiel A, Ozier O. *et al.* (2003) *Genome Res.*, 13, 2498–2504.

Thomas, M., Langley, B., Berry, C. *et al.* (2000) *J. Biol. Chem*., 275, 40235–40243.

van Oort, R.J., van Rooij, E., Bourajjaj, M. *et al.* (2006) *Circulation*, 114:298–308.

Vaquerizas, J.M., Kummerfeld, S.K., Teichmann, S.A. *et al.* (2009) *Nat. Rev. Genet.*, 10:252–263.

Yamada, T., Park, C.S., Mamonkin, M. *et al.* (2009) *Nat. Immunol*. 10:618–626.

Yusuf, I., Kharas, M.G., Chen, J. *et al.* (2008*) Int Immunol*. 20:671-681.