SCIENTIFIC REP⚙RTS

Functional and structural characterization of osteocytic MLO-Y4 cell proteins encoded by genes differentially expressed in response to mechanical signals in vitro

Fanchi Meng[1], Graeme F. Murray[2], Lukasz Kurgan[3] & Henry J. Donahue[2]

The anabolic response of bone to mechanical load is partially the result of osteocyte response to fluid flow-induced shear stress. Understanding signaling pathways activated in osteocytes exposed to fluid flow could identify novel signaling pathways involved in the response of bone to mechanical load. Bioinformatics allows for a unique perspective and provides key first steps in understanding these signaling pathways. We examined proteins encoded by genes differentially expressed in response to fluid flow in murine osteocytic MLO-Y4 cells. We considered structural and functional characteristics including putative intrinsic disorder, evolutionary conservation, interconnectedness in protein-protein interaction networks, and cellular localization. Our analysis suggests that proteins encoded by fluid flow activated genes have lower than expected conservation, are depleted in intrinsic disorder, maintain typical levels of connectivity for the murine proteome, and are found in the cytoplasm and extracellular space. Pathway analyses reveal that these proteins are associated with cellular response to stress, chemokine and cytokine activity, enzyme binding, and osteoclast differentiation. The lower than expected disorder of proteins encoded by flow activated genes suggests they are relatively specialized.

Mechanical fluctuations in the extracellular environment, such as interstitial fluid flow-induced shear stress, induce cell perturbations including membrane deformations, cytoskeletal restructuring, conformational changes in transmembrane proteins, changes in the glycol calyx and movement of cilia. These cell responses to mechanical signals are translated throughout the cell and can result in alterations in gene expression across many tissues[1].

In bone, osteocytes are well positioned to detect mechanical signals from loading and communicate them to downstream effector cells including osteoblasts, osteoclasts, and bone lining cells[2]. However, the signal transduction pathways involved in this process are only partially understood. A better understanding of signaling pathways activated by mechanical signals *in vitro* may lead to a better understanding of how bone adapts to load *in vivo*.

We have previously used gene microarrays, proteomics and RNA sequencing analysis (RNA-Seq) to identify increases in inflammatory C-X-C motif chemokines, including Ccl2, and the HIF-1α, IL-17, and AMPK signaling pathways induced by fluid flow[3,4]. However, these studies were limited as they did not examine functional and structural characteristics of proteins encoded by the corresponding differentially expressed genes and have focused only on genes with the highest signal. To address this we analyzed previously obtained RNA-seq data and gene micro array data[3,4] using several high-throughput computational methods to quantify evolutionary

[1]Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Canada. [2]Bone Engineering, Science and Technology (BEST) Laboratory, Department of Biomedical Engineering, Virginia Commonwealth University, Richmond, Virginia, United States of America. [3]Department of Computer Science, Virginia Commonwealth University, Richmond, Virginia, United States of America. Fanchi Meng and Graeme F. Murray contributed equally to this work. Correspondence and requests for materials should be addressed to L.K. (email: lkurgan@vcu.edu) or H.J.D. (email: hjdonahue@vcu.edu)

conservation, interconnectedness, and putative intrinsic disorder of the proteins encoded by all differentially expressed genes. We also functionally characterized these proteins using pathway analysis and gene ontology. We analyzed which of these characteristics are unique to proteins encoded by genes affected by fluid flow by comparing them to a generic set of proteins from the mouse proteome.

Conservation quantifies the amount of evolutionary variation in a protein sequence across species. When measured at the protein level it reflects the abundance of functionally and structurally important residues that are typically retained across species[5–7]. Interconnectedness measures the number of interactions between proteins in protein-protein interaction (PPI) networks, and can predict, along with conservation, the extent to which this protein is essential[8,9] and central to cellular processes[10]. Intrinsic disorder reflects the degree to which a protein lacks an ordered 3D structure under physiologic conditions[11,12]. While the levels of intrinsic disorder vary from protein to protein, disorder itself is ubiquitous. Recent estimates show that about 19% of amino acid residues are disordered in Eukaryotes[13] and between 6% and 17% of proteins encoded by various genomes are entirely disordered[14]. Intrinsic disorder is implicated in a diverse range of cellular functions including transcription[15,16], translation[17], protein-protein interactions[18,19], protein-RNA interactions[20,21] and cell signaling[19,22,23], to name a few. Functional characteristics that are specific to disorder include one-to-many binding, where one disordered region can interact with several structurally diverse partners; ability to fold upon binding; high accessibility to and regulation by post-translational modifications; and ability to implement cascade interactions[24]. At the same time, a high level of disorder is a substantial obstacle to solving protein structure[25–27] and to performing rational drug design that relies on knowledge of protein structure[28–30].

Gene ontology (GO) is a bioinformatics framework that defines biological processes, molecular functions and subcellular location of genes and proteins[31,32]. These annotations allow for the computational identification of functions and locations that are characteristic of a given group of genes or proteins. In a similar manner, pathway and network analysis can compare differentially expressed proteins and genes to known pathways to identify possible downstream effects.

We hypothesized that examining protein conservation, interconnectedness, intrinsic disorder, as well as pathway analysis, would provide insights to the function of osteocytic proteins encoded by genes differentially expressed in response to fluid flow-induced shear stress. This is the first study to examine evolutionary conservation, interconnectedness and putative intrinsic disorder in the context of mechanobiology.

Results

We comprehensively characterized several major structural and functional characteristics of osteocytic proteins encoded by genes differentially expressed due to exposure to fluid flow-induced shear stress. Figure 1 summarizes results focusing on evolutionary conservation, content of putative intrinsic disorder and interconnectedness. We compared the entire complement of proteins encoded by differentially expressed genes as well as proteins encoded by genes upregulated and downregulated against a generic set of murine proteins. Figure 1A reveals that proteins that correspond to the differentially expressed genes, and in particular proteins encoded by genes that are upregulated, have significantly lower than expected evolutionary conservation (p-value < 0.05). Proteins encoded by genes that are downregulated are also characterized by lower levels of conservation, however, the difference did not reach statistical significance when p-value considered significant was set at 0.05 (p-value $= 0.18$). Figure 1B reveals similar observations for the content of intrinsic disorder. That is, proteins encoded by differentially expressed and upregulated genes have significantly lower than expected disorder content (p-value < 0.05). Proteins encoded by downregulated genes also have lower disorder content, however, the difference did not reach statistical significance when p value considered significant was set at 0.05 (p-value $= 0.32$). The median disorder content across proteins encoded by differentially regulated genes equals 0.09 compared to the twice larger (0.18) content that is typical for murine proteins[13]. On the other hand, the degree of interconnectedness is not significantly different. Proteins encoded by differentially expressed genes, including both upregulated and downregulated subsets, have similar levels of PPIs (p-values > 0.6) compared to an overall set of murine proteins. The average number of PPIs in mouse listed in the metha resource equals 5.97[33] and is, as expected, virtually identical to the number that we report for our randomly chosen subset of murine proteins (6.03), and only slightly higher than for the proteins encoded by differentially expressed genes (5.63).

For GO analyses, we collected the most frequent term for each path in the GO hierarchy to reduce redundancy between the terms and to include the most frequent terms. Figure 2 depicts a list of subcellular locations (black bars), molecular functions (light gray bars) and biological processes (dark gray bars) that are significantly enriched in the proteins encoded by differentially expressed genes (p-value < 0.05 and at least 50% increase compared to the expected value to ensure that magnitude of the difference is substantial). Our genes of interest are primarily located in the cytoplasm and extracellular space. The GO terms that have high enrichment and high counts (second number inside square brackets next to the GO term name in Fig. 2) suggests that they are involved in chemokine (GO:0008009 and 0070098) and cytokine (GO:0034097) activities, response and regulation of response to stress (GO:0006950 and 0080134), and cellular response to chemical stimulus and oxygen-containing compound (GO:0070887 and 1901700). Annotations of molecular functions (light gray bars) also suggest that they take part in enzyme and G-protein coupled receptor binding.

We further focused on several key functions related to chemokine and cytokine processes and response and regulation of response to stress. We visualized the distributions of the evolutionary conservation, putative intrinsic disorder and interconnectedness for these proteins encoded by differentially expressed genes and compared them to a generic set of murine proteins (Fig. 3). Consistent with the results in Fig. 1, conservation (Fig. 3A) and amount of intrinsic disorder (Fig. 3B) are significantly lower than expected in these three sets of proteins that correspond to the differentially expressed genes (p-value < 0.05, except for the difference in conservation for proteins with chemokine activity where p-value $= 0.3$). Moreover, the numbers of PPIs are similar to a generic murine protein (p-value $= 0.99$).
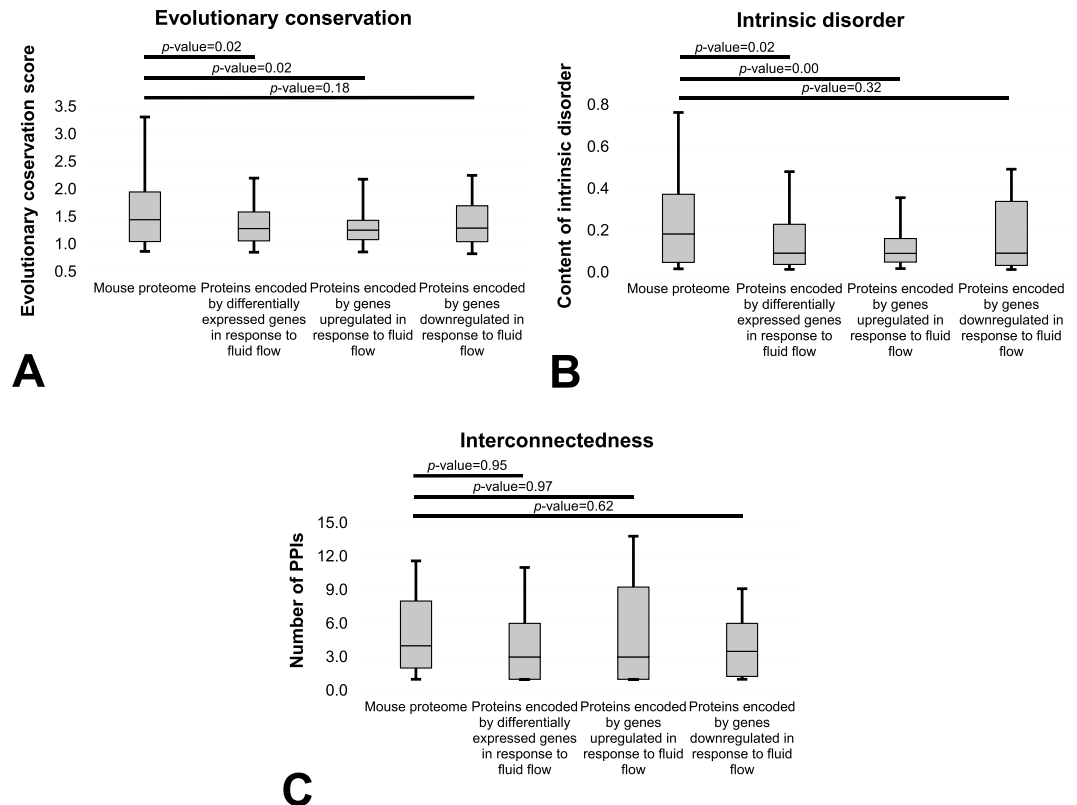
Figure 1. Comparison of the distrubutions of the evolutionary conservation quantified with the relative entropy scores (Panel A), content of the intrinsic disorder (Panel B) and the interconnectedness measured by the number of PPIs (Panel C) between the mouse proteins encoded by genes that are differentially expressed in response to fluid flow and a generic set of mouse proteins. The differentially expressed gene sets are also subdivided into upregulated and downregulated proteins. Distributions are represented with the box and whisker diagrams where the first quartile, second quartile (median) and third quartile are represented by the box and the whiskers correspond to the 10th and 90th centiles. Horizontal bars at the top of the figure summarize results of the analysis of statistical significance of the differences between the distribution for the two corresponding datasets. The significance was quantified with the two sample K-S test and the corresponding p-values are shown above the bars.

Ingenuity Pathway Analysis constructs causal networks from individual relationships detailed in curated literature[34]. Using these networks, downstream effects can be predicted based on our data. Casual network analysis of our RNA-seq data found a fluid flow-induced upregulation of proteins encoded by differentially expressed genes that are associated with the downregulation of osteoclast differentiation (Fig. 4). Shown in Fig. 4 are all the proteins that correspond to the differentially expressed genes in our dataset expected to influence osteoclast differentiation. NOS2, LPRC17 and PLK3R1 that are associated with inhibition of osteoclast differentiation were moderately upregulated; PTGS2 was also upregulated but would be predicted to increase osteoclast differentiation while TLR2 may effect osteoclast differentiation, IPA's database did not have enough power to generate a prediction. NOS2, TLR2, and PTGS2 all play a role in the regulation of the inflammatory response which is consistent with our gene ontology analysis which found increase of cytokines and chemokines.

Discussion

Our results suggest that proteins encoded by genes that are differentially expressed in response to fluid flow have lower than expected levels of sequence conservation and putative intrinsic disorder while maintaining typical levels of PPIs. Given their relatively low conservation, they are likely to carry out more specialized (non-essential) functions compared to the functions of proteins encoded by essential genes that have older evolutionary origins and are more conserved[34–36]. Proteins encoded by these essential genes are typically involved in basic cellular functions including gene expression, metabolism, morphogenesis, cell division, proliferation and differentiation, DNA replication, repair and transcription, and embryonic development and they are essential to the survival of the organism[34]. At the same time, the levels of interconnectedness of the proteins encoded by differentially expressed genes that are comparable to a typical mouse protein suggests that they are functionally important even though they may not be essential. The significant depletion in the putative intrinsic disorder suggests that they are highly structured and substantially more structured than a typical murine protein. The main implication of this observation is that structures of the proteins encoded by differentially expressed genes should be relatively easy to produce[25,37]. These structures can be used to decipher molecular level details of their functions[38] and they are necessary for virtual screening and design of novel therapeutics[39,40]. Taken together, our analyses suggest that genes involved
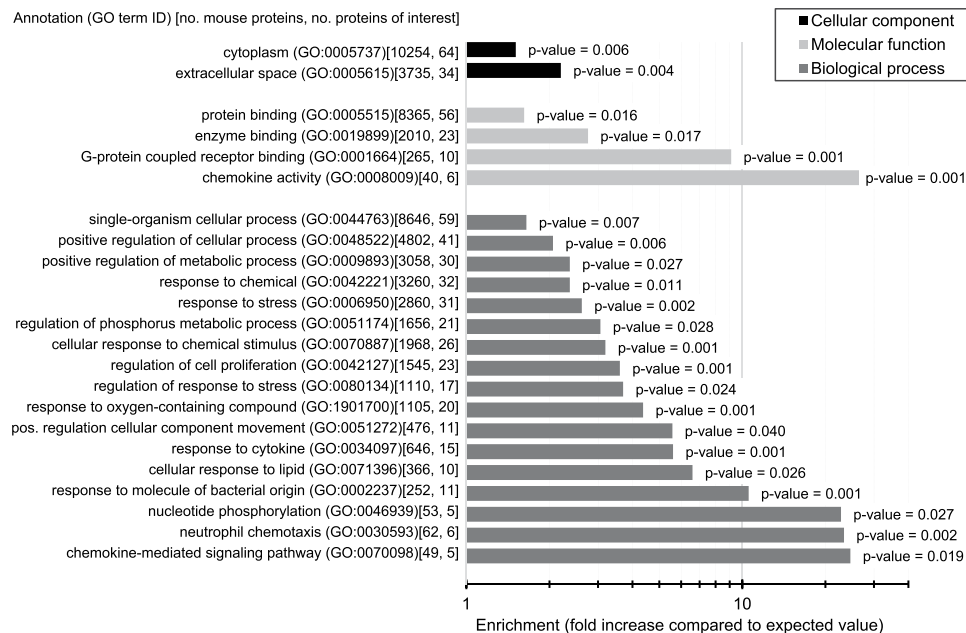
Figure 2. Summary of GO terms significantly enriched (*p*-value < 0.05) in the mouse proteins encoded by genes differentially expressed in response to fluid flow. The analysis was peformed using the PANTHER system[67] separately for each of the three classes of GO terms: molecular functions, biological processes and cellular component (subcellular location). Horizontal bars show the value of enrichment, defined as the fold increase when compared to the expected value measured on the mouse proteome, and the corresponding *p*-values are given next to the bars. The number of each GO term occurence in the mouse proteome and among the proteins encoded by differentially expressed genes is given inside the square brackets. We only consider GO terms that occur at least five time and for which the enrichment is greater than 150% and *p*-value < 0.05.

in mechanotransduction are non-essential, functionally important, and suitable for structure-based rational drug design. We note that the above observations rely on the putative intrinsic disorder and an incomplete PPI network, both of which may adversely affect accuracy of our analysis. However, we designed the analysis to reduce these effects. We employed a consensus of five predictors of disorder to minimize the prediction error and we used a recently released database that collects PPIs from a comprehensive set of five sources to provide the currently most complete network. Additionally, there are limitations to our fluid flow model that are shared by several *in vitro* models. These include analyses of only a single bone cell type, the two dimensional nature of our fluid flow apparatus and the fact that the cells are not in contact with a physiologically relevant extracellular matrix.

In this study we used IPA to assess RNA-seq data. We opted to assess RNA-seq data, rather than microarray data, because we previously showed that RNA-Seq yielded greater magnitude fold-changes in expression level than did microarrays, resulting in a broader overall dynamic range[3]. Furthermore, RNA-seq, compared to microarray was capable of detecting approximately 3 times the gene products of previously utilized microarrays. Our IPA analyses predicted fluid flow-induced changes in cytokines consistent with a down regulation of osteoclastogenesis, which one would expect as a result of exposure to anabolic mechanical signals. Additionally, consistent with our results, osteocytes have been shown to downregulate osteoclast differentiation through effects on TGF-β, an upstream inducer of IL-17, a factor that increases osteoclastogenesis[41–43]. On the other hand, our previous work showed fluid flow-induced upregulation of pathways dependent on IL-17[4] and another study showed that mechanical load upregulates TGF-β in mesenchymal stem cells[44]. These findings emphasize the complexity of the bone microenvironment and suggest the effects of cytokines on bone are highly context dependent. Ultimately, cytokines can have either anabolic or catabolic effects on bone which has complicated efforts at treatment of metabolic bone disease[45]. This complexity is illustrated by IL-17 which can be upregulated in both catabolic and anabolic states. Phase II trials of IL-17 inhibitors such as secukinumab for rheumatoid arthritis have yielded disappointing results[46]. Thus, there is a need for further research in this area to characterize the bone microenvironment an how it is affected by mechanical load, especially as regards the pleiotropic effect of cytokines in bone. The complexity of microenvironment also suggest the therapeutic target potential of mechanotransduction pathways because we have shown them to be less essential while still functionally important when compared to the mouse proteome.

In summary, this is the first study to quantify the structural and functional characteristics such as intrinsic disorder, evolutionary conservation, and interconnectedness of the proteins encoded by differentially expressed genes involved in mechanotransduction. These results suggest that the osteocytic signaling pathways activated by fluid flow are non-essential and populated by highly structured proteins. Furthermore, extracellular inflammatory mediators are involved in the downregulation of osteoclast differentiation. The well-defined tertiary structure of proteins encoded by differentially expressed genes in this pathway suggests that protein modeling could be successful in developing a better understanding of the signaling molecules involved in the response of bone to mechanical load.
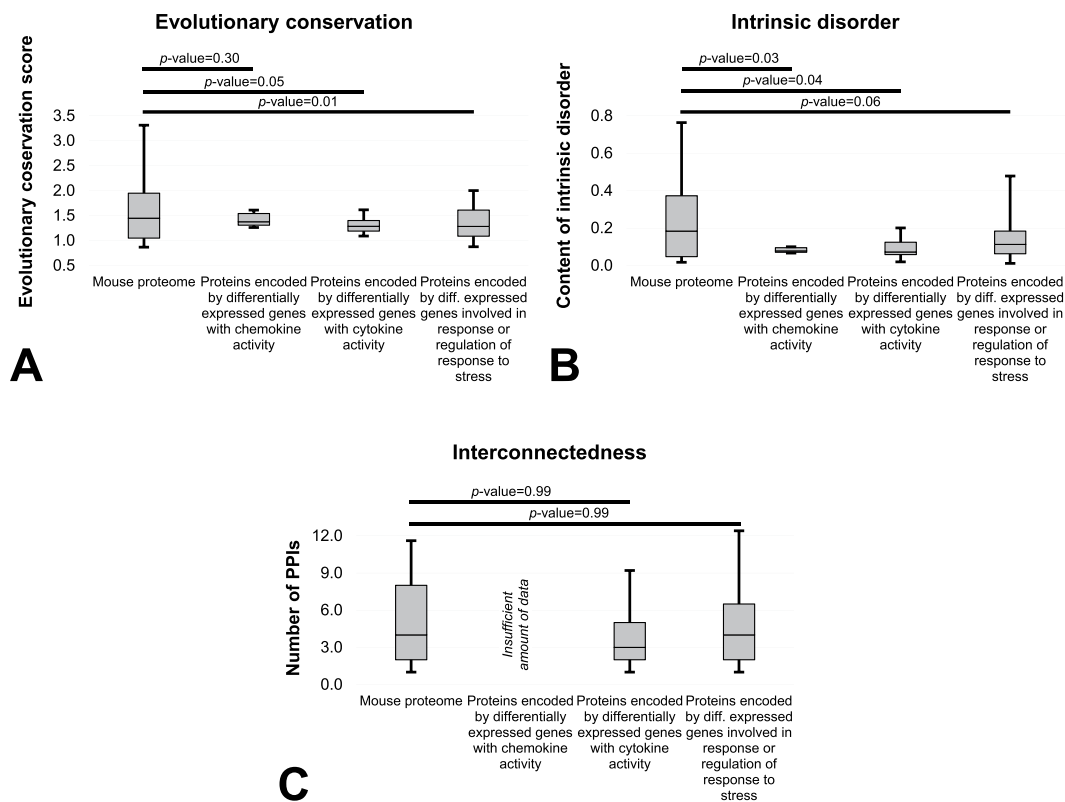
Figure 3. Comparison of the distrubutions of the evolutionary conservation quantified with the relative entropy scores (Panel A), content of the intrinsic disorder (Panel B) and the interconnectedness measured by the number of PPIs (Panel C) between functionally clustered subsets of the mouse proteins coded by genes that are differentially expressed in response to fluid flow (proteins of interest (POI)) and a generic set of mouse proteins. We considered three clusters that correspond to the proteins with chemokine activity (GO:0008009; 6 proteins), cytokine activity (GO:0034097; 15 proteins) and to those that are involved in the response of regulation of response to stress (GO:0006950 and 0080134; 35 proteins). Five of them were identified across the three clusters and additional eight are in common between the latter two clusters. Only three proteins encoded by differentially expressed genes with the cytokine activity had the PPI information and thus the analysis was not performed due to the low sample size. Distributions are represented with the box and whisker diagrams where the first quartile, second quartile (median) and third quartile are represented by the box and the whiskers correspond to the 10th and 90th centiles. Horizontal bars at the top of the figure summarize results of the analysis of statistical significance of the differences between the distribution for the two corresponding datasets. The significance was quantified with the two sample K-S test and the corresponding p-values are shown above the bars.
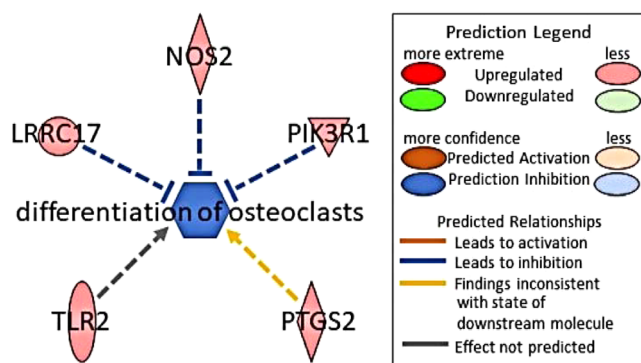
Figure 4. IPA predicted inhibition of the differentiation of osteoclasts with an activation Z score of -1.07 with a p-value < 0.001.

Methods

Collection of experimental data.   In the original study[3], murine osteocytic MLO-Y4 cells were cultured in normal growth medium (α-MEM [Invitrogen, Grand Island, NY] with 2.5% calf serum [Hyclone, Logan, UT], 2.5% fetal bovine serum [Lonza, Walkersville, MD], 1% Penicillin/Streptomycin) throughout all portions of the experiment. Cells were seeded 48 h prior to fluid flow on $75 \times 38 \times 1$ mm glass slides coated with $300\,\mu g/ml$ Type I Collagen (BD Biosciences, Bedford, MA) for 1 h and washed. Cell seeding density was 1.35×10^4 cells/cm^2 so that upon flow exposure, cells were roughly 60% confluent and interconnected by dendritic processes. Cells were then exposed to 2 hours of sinusoidally oscillating fluid flow in parallel plate flow chambers, inducing 1 Pa (10 dynes/cm^2) shear stress at 1 Hz. Paired sham controls were maintained in identical, static chambers. Triplicates of both flowed and static cells were collected following post-incubation in fresh medium for 2 hours.

Total RNA was isolated using Qiagen RNeasy Mini Kits, deep sequenced using Illumina HiSeq 2500, and processed computationally to identify genes that are differentially expressed between the samples exposed to the fluid flow and the controls[4]. Similar and complementary analysis was performed using DNA microarrays to identify further gene products that are characterized by differential abundance[3]. In both cases only the genes characterized by significant differences (p-values < 0.05) were selected. These data are available in the Gene Expression Omnibus with accession number GSE70667 for the RNA-Seq and GSE42874 for the microarrays.

Dataset.   Our prior analyses revealed two sets of 55 differentially expressed genes based on the RNA-Seq and microarray experiments. After removing duplicates of the 6 genes that were in common between the two sets, we mapped 99 genes to their UniGene identifiers. The remaining 5 genes could not be mapped to these identifiers. Next, we searched the UniProt resource[47,48] to map the UniGene identifiers to the corresponding mouse proteins. Once we removed duplicate proteins and protein fragments, we extracted 103 proteins encoded by the 99 differentially expressed genes that we were able to map into unique UniProt accession numbers. Among the 103 proteins encoded by the differentially expressed genes, 45 and 58 that are significantly upregulated or downregulated by fluid flow, respectively.

Analysis.   The 103 proteins encoded by differentially expressed genes in response to fluid flow were subjected to a comprehensive bioinformatics analysis. We contrasted their structural and functional characteristics against the reference mouse proteome collected from UniProt (C57BL/6 J strain, proteome ID: UP000000589). Since the amount of intrinsic disorder in eukaryotic organisms was shown to depend on the protein length[49], we randomly selected a matching number of mouse proteins to obtain the same distribution of lengths when compared to our 103 proteins encoded by differentially expressed genes. This ensures that our comparison of intrinsic disorder accommodates sequence-length bias; the same correction was made in several recent studies[50,51]. Equalization of sample size (103 proteins encoded by differentially expressed genes plus 103 proteins sampled from the reference proteome) also ensures that these data are suitable for robust statistical analysis. This is in contrast to using the whole mouse proteome of over 50 thousand proteins, which would result in over two orders of magnitude difference in sample size relative to proteins differentially expressed in response to fluid flow. The two sets have virtually identical distributions of protein length; p-value $= 1$ based on the two-sample Kolmogorov-Smirnov test. The difference in their median chain length, which is 412 and 416 for the random set of mouse proteins and our proteins encoded by differentially expressed genes, respectively, is not significant; p-value $= 0.97$ based on the Mann-Whitney test.

We estimated evolutionary conservation from multiple alignment profiles generated with HHblits against the *UniProt-20* database[52]. These profiles were used to compute relative entropy from each amino acid[53,54]. An average per-amino acid conservation was used to quantify conservation for each proteins encoded by differentially expressed genes. We collected the mouse PPI network from the *mentha* resource that integrates data from five manually curated source databases[33]. With that integrative approach, the mentha resource offers arguably the most complete currently available PPI network in mouse. The degree of interconnectedness was quantified by its number of PPIs. We annotated putative intrinsic disorder using a majority-vote based consensus of five predictors: three versions of the Espritz method[55] and two versions of the IUPred method[56]; These methods are characterized by complementary designs and were empirically demonstrated to offer strong predictive quality for the prediction of intrinsic disorder[57–59]. We used the consensus to minimize the prediction error, which is in line with an observation that this results in a better predictive quality when compared to the use of individual predictors[60,61]. Such consensuses are also implemented in the current databases of putative disorder: MobiDB[62,63] and D^{2}P^2[64], and were utilized in many other studies[13,17,21,50,51,65,66]. We quantified amount of disorder with the disorder content, which is defined as a fraction of disordered amino acids in a given sequence. We measured statistical significance of differences in distributions of conservation, disorder and interconnectedness values between our proteins of interest and the mouse proteome using a two-sample Kolmogorov-Smirnov test (K-S test) and assumed that a given difference is significant at p-values < 0.05.

We performed functional analysis using GO with the PANTHER system[67]. PANTHER performs enrichment analysis of GO terms associated with our proteins encoded by differentially expressed genes against the expected frequency of these terms in the mouse proteome[68]. We executed this analysis separately for each of the three classes of GO terms: molecular functions, biological processes and cellular component (subcellular location). The pathway and causal network analyses was completed using the Ingenuity Pathway Analysis (IPA) tool[69].

References

1. Paluch, E. K. *et al.* Mechanotransduction: use the force(s). *BMC Biol* **13**, 47, https://doi.org/10.1186/s12915-015-0150-4 (2015).
2. Bonewald, L. F. The amazing osteocyte. *J Bone Miner Res* **26**, 229–238, https://doi.org/10.1002/jbmr.320 (2011).
3. Govey, P. M. *et al.* Integrative transcriptomic and proteomic analysis of osteocytic cells exposed to fluid flow reveals novel mechanosensitive signaling pathways. *J Biomech* **47**, 1838–1845, https://doi.org/10.1016/j.jbiomech.2014.03.022 (2014).
4. Govey, P. M., Kawasawa, Y. I. & Donahue, H. J. Mapping the osteocytic cell response to fluid flow using RNA-Seq. *Journal of biomechanics* **48**, 4327–4332 (2015).

5. Zhang, T. *et al*. Analysis and prediction of RNA-binding residues using sequence, evolutionary conservation, and predicted secondary structure and solvent accessibility. *Curr Protein Pept Sci* **11**, 609–628 (2010).

6. Chakrabarti, S. & Lanczycki, C. J. Analysis and prediction of functionally important sites in proteins. *Protein science: a publication of the Protein Society* **16**, 4–13, https://doi.org/10.1110/ps.062506407 (2007).

7. Fang, C., Noguchi, T. & Yamana, H. Analysis of evolutionary conservation patterns and their influence on identifying protein functional sites. *J Bioinform Comput Biol* **12**, 1440003, https://doi.org/10.1142/S0219720014400034 (2014).

8. He, X. L. & Zhang, J. Z. Why do hubs tend to be essential in protein networks? *PLoS genetics* **2**, 826–834, https://doi.org/10.1371/journal.pgen.0020088 (2006).

9. Zotenko, E., Mestre, J., O'Leary, D. P. & Przytycka, T. M. Why Do Hubs in the Yeast Protein Interaction Network Tend To Be Essential: Reexamining the Connection between the Network Topology and Essentiality. *Plos Comput Biol* **4**, https://doi.org/10.1371/journal.pcbi.1000140 (2008).

10. Han, J. D. *et al*. Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature* **430**, 88–93, https://doi.org/10.1038/nature02555 (2004).

11. Dunker, A. K. *et al*. What's in a name? Why these proteins are intrinsically disordered. *Intrinsically Disordered Proteins* **1**, e24157, https://doi.org/10.4161/idp.24157 (2013).

12. van der Lee, R. *et al*. Classification of Intrinsically Disordered Regions and Proteins. *Chemical Reviews* **114**, 6589–6631, https://doi.org/10.1021/cr400525m (2014).

13. Peng, Z. *et al*. Exceptionally abundant exceptions: comprehensive characterization of intrinsic disorder in all domains of life. *Cell Mol Life Sci* **72**, 137–151, https://doi.org/10.1007/s00018-014-1661-9 (2015).

14. Tompa, P. Intrinsically unstructured proteins. *Trends in Biochemical Sciences* **27**, 527–533, https://doi.org/10.1016/S0968-0004(02)02169-2 (2002).

15. Liu, J. *et al*. Intrinsic disorder in transcription factors. *Biochemistry* **45**, 6873–6888, https://doi.org/10.1021/bi0602718 (2006).

16. Fuxreiter, M. *et al*. Malleable machines take shape in eukaryotic transcriptional regulation. *Nat Chem Biol* **4**, 728–737, https://doi.org/10.1038/nchembio.127 (2008).

17. Peng, Z. *et al*. A creature with a hundred waggly tails: intrinsically disordered proteins in the ribosome. *Cell Mol Life Sci* **71**, 1477–1504, https://doi.org/10.1007/s00018-013-1446-6 (2014).

18. Fuxreiter, M. *et al*. Disordered proteinaceous machines. *Chem Rev* **114**, 6806–6843, https://doi.org/10.1021/cr4007329 (2014).

19. Dunker, A. K., Cortese, M. S., Romero, P., Iakoucheva, L. M. & Uversky, V. N. Flexible nets. *The roles of intrinsic disorder in protein interaction networks*. *FEBS J* **272**, 5129–5148, https://doi.org/10.1111/j.1742-4658.2005.04948.x (2005).

20. Varadi, M., Zsolyomi, F., Guharoy, M. & Tompa, P. Functional Advantages of Conserved Intrinsic Disorder in RNA-Binding Proteins. *PloS one* **10**, e0139731, https://doi.org/10.1371/journal.pone.0139731 (2015).

21. Wang, C., Uversky, V. N. & Kurgan, L. Disordered nucleiome: Abundance of intrinsic disorder in the DNA- and RNA-binding proteins in 1121 species from Eukaryota, Bacteria and Archaea. *Proteomics* **16**, 1486–1498, https://doi.org/10.1002/pmic.201500177 (2016).

22. Xie, H. *et al*. Functional anthology of intrinsic disorder. 1. *Biological processes and functions of proteins with long disordered regions*. *J Proteome Res* **6**, 1882–1898, https://doi.org/10.1021/pr060392u (2007).

23. Galea, C. A., Wang, Y., Sivakolundu, S. G. & Kriwacki, R. W. Regulation of cell division by intrinsically unstructured proteins: intrinsic flexibility, modularity, and signaling conduits. *Biochemistry* **47**, 7598–7609, https://doi.org/10.1021/bi8006803 (2008).

24. Uversky, V. N. A decade and a half of protein intrinsic disorder: biology still waits for physics. *Protein science: a publication of the Protein Society* **22**, 693–724, https://doi.org/10.1002/pro.2261 (2013).

25. Oldfield, C. J. *et al*. Utilization of protein intrinsic disorder knowledge in structural proteomics. *Biochim. Biophys. Acta* **1834**, 487–498, https://doi.org/10.1016/j.bbapap.2012.12.003 (2013).

26. Linding, R. *et al*. Protein disorder prediction: implications for structural proteomics. *Structure* **11**, 1453–1459 (2003).

27. Mizianty, M. J. & Kurgan, L. Sequence-based prediction of protein crystallization, purification and production propensity. *Bioinformatics (Oxford, England)* **27**, i24–33, https://doi.org/10.1093/bioinformatics/btr229 (2011).

28. Jazayeri, A., Dias, J. M. & Marshall, F. H. From G Protein-coupled Receptor Structure Resolution to Rational Drug Design. *J Biol Chem* **290**, 19489–19495, https://doi.org/10.1074/jbc.R115.668251 (2015).

29. Lounnas, V. *et al*. Current progress in Structure-Based Rational Drug Design marks a new mindset in drug discovery. *Comput Struct Biotechnol J* **5**, e201302011, https://doi.org/10.5936/csbj.201302011 (2013).

30. Gane, P. J. & Dean, P. M. Recent advances in structure-based rational drug design. *Curr Opin Struct Biol* **10**, 401–404 (2000).

31. Ashburner, M. *et al*. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25–29, https://doi.org/10.1038/75556 (2000).

32. Gene Ontology, C. Gene Ontology Consortium: going forward. *Nucleic Acids Res* **43**, D1049–1056, https://doi.org/10.1093/nar/gku1179 (2015).

33. Calderone, A., Castagnoli, L. & Cesareni, G. Mentha: a resource for browsing integrated protein-interaction networks. *Nature Methods* **10**, 690–691 (2013).

34. Kabir, M., Barradas, A., Tzotzos, G. T., Hentges, K. E. & Doig, A. J. Properties of genes essential for mouse development. *PLoS One* **12**, e0178273, https://doi.org/10.1371/journal.pone.0178273 (2017).

35. Jordan, I. K., Rogozin, I. B., Wolf, Y. I. & Koonin, E. V. Essential genes are more evolutionarily conserved than are nonessential genes in bacteria. *Genome Res*. **12**, 962–968, https://doi.org/10.1101/gr.87702 (2002). Article published online before print in May 2002.

36. Luo, H., Gao, F. & Lin, Y. Evolutionary conservation analysis between the essential and nonessential genes in bacterial genomes. *Sci. Rep*. **5**, 13210, https://doi.org/10.1038/srep13210 (2015).

37. Oldfield, C. J., Ulrich, E. L., Cheng, Y., Dunker, A. K. & Markley, J. L. Addressing the intrinsic disorder bottleneck in structural proteomics. *Proteins* **59**, 444–453, https://doi.org/10.1002/prot.20446 (2005).

38. Teichmann, S. A., Murzin, A. G. & Chothia, C. Determination of protein function, evolution and interactions by structural genomics. *Curr Opin Struct Biol* **11**, 354–363 (2001).

39. Waszkowycz, B. Structure-based approaches to drug design and virtual screening. *Current opinion in drug discovery & development* **5**, 407–413 (2002).

40. Rognan, D. Structure-Based Approaches to Target Fishing and Ligand Profiling. *Mol. Inform*. **29**, 176–187, https://doi.org/10.1002/minf.200900081 (2010).

41. Okamoto, K. *et al*. Osteoimmunology: The Conceptual Framework Unifying the Immune and Skeletal Systems. *Physiol Rev* **97**, 1295–1349, https://doi.org/10.1152/physrev.00036.2016 (2017).

42. Gaffen, S. L. The role of interleukin-17 in the pathogenesis of rheumatoid arthritis. *Current rheumatology reports* **11**, 365–370 (2009).

43. Heino, T. J., Hentunen, T. A. & Vaananen, H. K. Osteocytes inhibit osteoclastic bone resorption through transforming growth factor-beta: enhancement by estrogen. *J Cell Biochem* **85**, 185–197 (2002).

44. Gardner, O. F., Fahy, N., Alini, M. & Stoddart, M. J. Joint mimicking mechanical load activates TGFbeta1 in fibrin-poly(ester-urethane) scaffolds seeded with mesenchymal stem cells. *J Tissue Eng Regen Med*. https://doi.org/10.1002/term.2210 (2016).

45. Jung, S. M., Kim, K. W., Yang, C. W., Park, S. H. & Ju, J. H. Cytokine-mediated bone destruction in rheumatoid arthritis. *Journal of immunology research* **2014**, 263625, https://doi.org/10.1155/2014/263625 (2014).

46. Koenders, M. I. & van den Berg, W. B. Secukinumab for rheumatology: development and its potential place in therapy. *Drug design, development and therapy* **10**, 2069–2080, https://doi.org/10.2147/dddt.s105263 (2016).

47. Apweiler, R. *et al.* UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res* **32**, D115–119, https://doi.org/10.1093/nar/gkh131 (2004).
48. UniProt, C. UniProt: a hub for protein information. *Nucleic Acids Res* **43**, D204–212, https://doi.org/10.1093/nar/gku989 (2015).
49. Howell, M. *et al.* Not That Rigid Midgets and Not So Flexible Giants: On the Abundance and Roles of Intrinsic Disorder in Short and Long Proteins. *J Biol Syst* **20**, 471–511, https://doi.org/10.1142/S0218339012400086 (2012).
50. Meng, F., Na, I., Kurgan, L. & Uversky, V. N. Compartmentalization and Functionality of Nuclear Disorder: Intrinsic Disorder and Protein-Protein Interactions in Intra-Nuclear Compartments. *Int J Mol Sci* **17**, https://doi.org/10.3390/ijms17010024 (2016).
51. Na, I., Meng, F., Kurgan, L. & Uversky, V. N. Autophagy-related intrinsically disordered proteins in intra-nuclear compartments. *Mol Biosyst* **12**, 2798–2817, https://doi.org/10.1039/c6mb00069j (2016).
52. Remmert, M., Biegert, A., Hauser, A. & Söding, J. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nature Methods* **9**, 173–175 (2012).
53. Wang, K. & Samudrala, R. Incorporating background frequency improves entropy-based residue conservation measures. *BMC bioinformatics* **7**, 385 (2006).
54. Wang, K. & Samudrala, R. Automated functional classification of experimental and predicted protein structures. *BMC bioinformatics* **7**, 278 (2006).
55. Walsh, I., Martin, A. J., Di Domenico, T. & Tosatto, S. C. ESpritz: accurate and fast prediction of protein disorder. *Bioinformatics (Oxford, England)* **28**, 503–509, https://doi.org/10.1093/bioinformatics/btr682 (2012).
56. Dosztanyi, Z., Csizmok, V., Tompa, P. & Simon, I. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics (Oxford, England)* **21**, 3433–3434, https://doi.org/10.1093/bioinformatics/bti541 (2005).
57. Peng, Z. L. & Kurgan, L. Comprehensive comparative assessment of in-silico predictors of disordered regions. *Curr Protein Pept Sci* **13**, 6–18 (2012).
58. Walsh, I. *et al.* Comprehensive large-scale assessment of intrinsic protein disorder. *Bioinformatics (Oxford, England)* **31**, 201–208, https://doi.org/10.1093/bioinformatics/btu625 (2015).
59. Meng, F. & Uversky, V. N. & Kurgan, L. Comprehensive review of methods for prediction of intrinsic disorder and its molecular functions. *Cell Mol Life Sci.* https://doi.org/10.1007/s00018-017-2555-4 (2017).
60. Peng, Z. & Kurgan, L. On the complementarity of the consensus-based disorder prediction. *Pac Symp Biocomput*, 176–187 (2012).
61. Fan, X. & Kurgan, L. Accurate prediction of disorder in protein chains with a comprehensive and empirically designed consensus. *Journal of biomolecular structure & dynamics* **32**, 448–464, https://doi.org/10.1080/07391102.2013.775969 (2014).
62. Di Domenico, T., Walsh, I., Martin, A. J. M. & Tosatto, S. C. E. MobiDB: a comprehensive database of intrinsic protein disorder annotations. *Bioinformatics (Oxford, England)* **28**, 2080–2081, https://doi.org/10.1093/bioinformatics/bts327 (2012).
63. Potenza, E., Di Domenico, T., Walsh, I. & Tosatto, S. C. MobiDB 2.0: an improved database of intrinsically disordered and mobile proteins. *Nucleic Acids Res* **43**, D315–320, https://doi.org/10.1093/nar/gku982 (2015).
64. Oates, M. E. *et al.* D(2)P(2): database of disordered protein predictions. *Nucleic Acids Res* **41**, D508–516, https://doi.org/10.1093/nar/gks1226 (2013).
65. Hu, G., Wu, Z., Wang, K., Uversky, V. N. & Kurgan, L. Untapped Potential of Disordered Proteins in Current Druggable Human Proteome. *Curr Drug Targets* **17**, 1198–1205 (2016).
66. Peng, Z., Uversky, V. N. & Kurgan, L. Genes Encoding Intrinsic Disorder in Eukaryota Have High GC Content. *Intrinsically Disordered Proteins* **4**, e1262225 (2016).
67. Mi, H., Muruganujan, A. & Thomas, P. D. PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Res* **41**, D377–386, https://doi.org/10.1093/nar/gks1118 (2013).
68. Mi, H., Muruganujan, A., Casagrande, J. T. & Thomas, P. D. Large-scale gene function analysis with the PANTHER classification system. *Nat Protoc* **8**, 1551–1566, https://doi.org/10.1038/nprot.2013.092 (2013).
69. Kramer, A., Green, J., Pollard, J. Jr & Tugendreich, S. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics (Oxford, England)* **30**, 523–530, https://doi.org/10.1093/bioinformatics/btt703 (2014).

Acknowledgements

Author Contributions

F.M. collected the data, performed analysis, summarized the results, and contributed to writing the manuscript, G.M. performed data analyses and contributed to writing the manuscript, L.K. conceptualized the study, performed analysis, summarized and analyzed results, and wrote the manuscript, H.D. collected original experimental data, conceptualized the study and wrote the manuscript.

Additional Information