

## Mapping arginine methylation in the human body and cardiac disease

Donatus O. Onwuli,<sup>1</sup> Laura Rigau-Roca,<sup>2</sup> Chris Cawthorne,<sup>1</sup> Pedro Beltran-Alvarez<sup>1,\*</sup>

<sup>1</sup> School of Life Sciences, University of Hull, Hull, United Kingdom.

<sup>2</sup> St Mary's College, Hull, United Kingdom

\* Corresponding author: Dr. Pedro Beltran-Alvarez. School of Life Sciences, Hardy Building,

University of Hull, Cottingham Road, Hull HU6 7RX, United Kingdom. Tel No. +441482466624.

Fax: +441482466511. E-mail: [p.beltran-alvarez@hull.ac.uk](mailto:p.beltran-alvarez@hull.ac.uk)

**Keywords:** arginine methylation, data mining, heart disease, post-translational modifications, tissue

**Abbreviations:** Arginine methylation (ArgMe), Post-Translational Modification (PTM), Protein Arginine Methyl Transferases (PRMTs), *S*-adenosyl-L-methionine (SAM).

**Number of words including references and legends:** 5146

**Abstract:**

**Purpose**

Arginine methylation (ArgMe) is one of the most ubiquitous post-translational modifications, and hundreds of proteins undergo ArgMe in *e.g.* brain. However, the scope of ArgMe in many tissues, including the heart, is currently under explored. Here, we aimed to 1) identify proteins undergoing ArgMe in human organs, and 2) expose the relevance of ArgMe in cardiac disease.

Received: 11/08/2016; Revised: 20/08/2016; Accepted: 03/09/2016

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1002/prca.201600106](#).

This article is protected by copyright. All rights reserved.

## **Experimental design**

We used publicly available proteomic data to search for ArgMe in 13 human tissues. We used glucose to induce H9c2 cardiac-like cell hypertrophy.

## **Results**

Our results show that ArgMe is mainly tissue-specific; nevertheless, we suggest an embryonic origin of core ArgMe events. In the heart, we found 103 mostly novel ArgMe sites in 58 non-histone proteins. We provide compelling evidence that cardiac protein ArgMe is relevant to cardiomyocyte ontology, and important for proper cardiac function. This is highlighted by the fact that genetic mutations affecting methylated arginine positions are often associated with cardiac disease, including hypertrophic cardiomyopathy. We provide pilot experimental data suggesting significant changes in ArgMe profiles of H9c2 cells upon induction of cell hypertrophy using glucose.

## **Conclusions and clinical relevance**

Our work calls for in-depth investigation of ArgMe in normal and diseased tissues, using methods including clinical proteomics.

## **Statement of Clinical Relevance**

Arginine methylation (ArgMe) is the subject of intense research in the context of cancer, which has included the development and testing of tens of ArgMe inhibitors in pre-clinical studies over the last decade. Here, we propose to further the research of ArgMe in the setting of cardiac disease. We provide evidence supporting the notion that ArgMe plays key roles in cardiac tissue morphogenesis and function. Our pilot data suggest that cardiac protein ArgMe profiles change significantly upon induction of cardiac-like cell hypertrophy using high glucose. Based on these results, we believe that researchers with an interest in cardiology and clinical proteomics could consider research in ArgMe as a new, and clinically relevant, research opportunity, given 1) the growing number of (bio)chemical and analytical tools to study ArgMe, 2) the increasing incidence of diabetic cardiomyopathy, and 3)

the need for new strategies for prevention, diagnosis, and treatment of cardiac hypertrophy, a disease with no current cure.

## 1. Introduction

Arginine methylation (ArgMe) is one of the *ca.* 200 known protein post-translational modifications (PTMs) [1], whereby one (monoArgMe) or two (diArgMe) methyl groups are transferred from *S*-adenosyl-L-methionine (SAM) to target arginine residues within proteins. ArgMe is catalysed by Protein Arginine Methyl Transferases (PRMTs), of which there are at least 9 distinct enzymes encoded in the human genome. These enzymes are named PRMT1-9, and are classified into three different types. While all PRMTs catalyse monoArgMe, Type I PRMTs including PRMT1, -3, -4, -6, and -8, additionally lead to asymmetric diArgMe ( $\omega$ -N<sup>G</sup>,N<sup>G</sup>-dimethylarginine), while Type II PRMTs such as PRMT5 and -9 symmetrically dimethylate target Arg residues leading to  $\omega$ -N<sup>G</sup>,N<sup>G</sup>-dimethylarginine. PRMT7 is a Type III PRMT and generates monoArgMe only. MonoArgMe, asymmetric and symmetric diArgMe differently increase the hydrophobicity and size of the residue, and change the geometry of hydrogen bonding [2].

Recent studies have revealed 500-1000 ArgMe sites in HCT116 cells (a colon cancer cell line), primary T cells, and mouse brain [3, 4]. These figures make ArgMe one of the top 10 most ubiquitous PTMs [5], however, ArgMe has not been comprehensively investigated in many tissues. For instance, only a handful of non-histone proteins, including Desmoplakin, GATA4, Na<sub>v</sub>1.5 and PGC-1 $\alpha$ , are known to undergo ArgMe in the heart [6-11].

PRMTs have been described in animal cells, yeast, and plant cells [12], and homologs of the two major PRMTs (*i.e.* PRMT1 and -5) have been reported in many eukaryotes [13]. Many PRMTs are essential during animal embryogenesis; for example PRMT1 is critical for mouse embryo viability [14, 15], and zebrafish embryo development [16]. PRMT3 null mice show retarded growth during gestation [17], PRMT4 (also known as CARM1) null mice die perinatally [18], and loss of PRMT5 is embryonic lethal [19]. PRMT2, -6, -7, and -8 knockout mice are viable, although specific

abnormalities have been reported [20-23]. Despite these and other landmark studies, we are still lacking an understanding of the conservation of ArgMe along tissues that share common embryonic origin (*e.g.* that derive from a given germ layer).

Recent high-resolution proteomic investigations by the group of Akhilesh Pandey have provided the proteomes of several human tissues, including that of the heart [24]. Importantly, the raw data as acquired by mass spectrometry are available in public data repositories. ArgMe is considered a stable PTM [2, 25], and therefore published proteomic data are amenable to further interrogation for ArgMe. To uncover the scope of ArgMe in the human body, we searched the data of Pandey (13 human tissues) for ArgMe. To test whether ArgMe events are mostly organ-specific or, on the other hand, have a common embryonic origin, we compared proteins undergoing ArgMe in tissues derived from the 3 germ layers (endoderm, mesoderm and ectoderm). We further set ourselves to uncover the relevance of ArgMe in the physiology of the heart by analysing proteomic and genetic data as well as by using H9c2 cardiac-like cells as an experimental model of cardiac cell hypertrophy.

## 2. Materials and methods

### 2.1 Bioinformatics

We downloaded raw data from the PRIDE database, Project PXD000561 by Pandey *et al.* [24]. For each tissue, all available files (typically 48-124 files, 500-1000 Mb each) were analysed using MaxQuant (v. 1.5.3.28) to quest the raw data for protein ArgMe. We set Arg/Lys mono- and dimethylation, as well as Met oxidation, and acetylation of protein N-terminus, as variable protein modifications. This enabled database searching for Arg residues taking into account that they may or may not be methylated. Cys carbamidomethylation was set as a fixed modification. Results were searched against SwissProt Human database (April, 2015). The rest of MaxQuant parameters was essentially as default. Allowed peptide length range was 8-25 amino acids. False discovery rate was set at 0.01, and minimum score for modified peptides was set at 40. Peptide mass tolerance was 4.5 ppm for parent ion spectra and 20 ppm for MS/MS spectra. A maximum of two missed cleavages was allowed.

For the analysis of cardiac proteins undergoing ArgMe in fetal and adult heart, we worked with 117 raw files for adult heart, corresponding to the pooling of 15 adult, histologically normal, cardiac samples; and 104 files for fetal heart, corresponding to the pooling of 12 fetal, histologically normal, cardiac samples. MaxQuant analysis was done as before.

ArgMe sites in the different tissues were compared using online Venn diagrams makers (<http://bioinformatics.psb.ugent.be/webtools/Venn/>). Gene names and protein IDs were as in ENSEMBL/ UNIPROT. Gene Ontology (GO) term enrichment analysis was done using GOrilla [26, 27], with two sets of data (target and background lists of proteins). Target proteins were proteins undergoing ArgMe in a tissue / group of tissues, and background was the proteome of the corresponding tissue (as identified by MaxQuant).

## **2.2 H9c2 cells and induction of H9c2 cell hypertrophy**

H9c2 embryonic cardiac myoblast cells were purchased from ATCC UK (CRL-1446), and maintained in low glucose Dulbecco Modified Eagle's Medium (Gibco, Gaithersburg, MD) supplemented with 10% fetal bovine serum (Gibco) in a humidified incubator at 5% CO<sub>2</sub> and 37 °C. For subculturing, cells were passaged every 2-3 days at 80% confluence. Induction of H9c2 hypertrophy using glucose was done following recently published protocols [28, 29]. Briefly, the day after seeding, H9c2 cell hypertrophy was induced by adding 30 mM glucose (final concentration, from 1 M sterile glucose stock) to the growth medium. Cultures were maintained for 72 h under these conditions, and then processed for cell size imaging and Western blot (see below). In parallel, H9c2 control cells were cultured without the addition of extra glucose.

## **2.3 Measurement of H9c2 cell size**

Experiments were started by seeding 1 x 10<sup>5</sup> H9c2 cells on cover slips in 6-well cell culture plates, and H9c2 cell hypertrophy was induced using glucose as above. After 72 hours treatment, cells were thoroughly washed, permeabilised using 1% Triton X-100, and incubated with 1% Rhodamine Phalloidin [#R415, ThermoFisher, Grand Island, NY] for 30 minutes at room temperature to visualise

F-actin by red/orange fluorescence. Slides were examined by fluorescence microscopy, and cell size was quantified using Image J [30]. Experiments were done in duplicate, values are reported as mean  $\pm$  SD, and t-student test was used to assign statistical significance.

#### **2.4 Detection of H9c2 protein ArgMe using Western blot**

Experiments were started by seeding  $1.6 \times 10^6$  H9c2 cells in T75 cell culture flasks, and H9c2 cell hypertrophy was induced using glucose as above. After 72 hours treatment, cells were thoroughly washed in PBS, harvested, and pelleted. Cell pellets were lysed in 100  $\mu$ l Laemmli buffer, and samples were briefly sonicated to reduce viscosity. Total protein lysates were then resolved on 10% acrylamide SDS-PAGE gels, and proteins transferred to nitrocellulose membranes (GE Healthcare Life Sciences, Buckinghamshire, UK) by Western blot. Membranes were probed using specific antibodies targeting mono- and diArgMe antibodies (#8015 and #13522, Cell Signalling Technology, Danvers, MA), as well as  $\beta$ -actin and GAPDH (both from Abcam, Cambridge, UK) as protein loading controls, and developed using ECL reagents (#WBLUF0100, Millipore, Hertfordshire, UK). Experiments were done in duplicate.

### **3. Results**

#### **3.1 Most ArgMe sites are tissue specific, although core ArgMe events seem to have an embryonic origin**

To define the “arginine methylome” of a number of human tissues, we interrogated the data previously published by the group of Pandey [24], for protein ArgMe in 13 human tissues (see flow chart in Figure 1A). We selected tissues derived from the 3 germ layers including colon, lung, pancreas, rectum, urinary bladder (endoderm), kidney, heart, ovary, prostate, testis (mesoderm), frontal cortex, retina, and spinal cord (ectoderm). Using MaxQuant, we generated a list of proteins undergoing mono- and diArgMe in each tissue (Supp. Tables 1 to 13). On average, 224 proteins (*ca.* 337 sites) undergoing ArgMe were detected per tissue at a false discovery rate of 1%.

Accepted Article

Approximately 53% and 58% mono- and diArgMe sites, respectively, were tissue specific. We defined “core” ArgMe events as those ArgMe sites conserved in at least 4 tissues in endoderm- or mesoderm-derived tissues, or throughout the 3 ectoderm-derived tissues. On average, core ArgMe events represented 7% and 13% of the total number of mono- and diArgMe sites per tissue, respectively (Table 1).

We then compared core ArgMe events among germ layers. We identified 5 mono- and 3 diArgMe sites conserved in all germ layers. We classified some ArgMe sites common to endoderm-and-mesoderm, or mesoderm-and-ectoderm, but we did not find ArgMe sites common to endoderm-and-ectoderm derived tissues (Figures 2A and 2B, and Table 2). To validate our results, we assessed the novelty of our 56 core ArgMe events, and found that 25 had previously been reported [31], (Table 2). It should be mentioned at this point that during the manual curation step a range of potential contaminants, including hemoglobins, keratins and immunoglobulins, were found as undergoing core ArgMe, but excluded from our analyses.

### **3.2 One hundred and three ArgMe sites in 58 non-histone proteins are conserved in fetal and adult heart**

Based on our results above, we hypothesized that ArgMe sites essential to tissue ontology must be conserved in fetal and adult tissues. We decided to test this hypothesis using the heart as a model system. First, we analysed protein expression and ArgMe in fetal (Supp. Table 14) compared to adult samples (Figure 1B, and Figures 2C and 2D). We found that 103 ArgMe sites in 58 non-histone proteins were conserved in fetal and adult heart (Supp. Table 15). Of note, 91 out of these 103 ArgMe sites were novel (Supp. Table 15), whereas 12 sites had previously been observed in other cells / tissues [31].

### 3.3 GO terms essential to cardiomyocytes are enriched in proteins undergoing ArgMe in the heart

Next, we addressed the question of whether cardiac protein ArgMe is relevant to cardiomyocyte physiology. We searched for enrichment in GO terms in our subset of 62 methylated cardiac proteins (including 58 found here, as well as Desmoplakin, GATA4, Na<sub>v</sub>1.5 and PGC-1 $\alpha$ ). We used these 62 methylated proteins as target set, and the 3124 proteins expressed in both fetal and adult heart as background set. We run the search using three ontologies, *i.e.* function, component and process. Importantly, GO terms related to cardiac muscle contraction, cardiomyocyte cytoskeleton and cardiac tissue morphogenesis were clearly enriched in methylated proteins (Supp. Table 16).

### 3.4 Damaging mutations have been reported in many ArgMe sites of cardiac proteins

We searched for human mutations at Arg positions found to be methylated in cardiac proteins. We queried the public version of the Human Gene Mutation Database [32], and the Exome Aggregation Consortium (ExAC) browser (Cambridge, MA, <http://exac.broadinstitute.org>), both last accessed April 2016. Titin was excluded from this analysis as many mutations in titin are currently of unknown significance [33]. In more than half of the cases (64 out of 109 methylation sites), we found mutations precisely at the would-have-been-methylated Arg residues. All mutations were rare (allele frequency < 0.001, as available in the ExAC browser). The vast majority of the mutations (50 out of 64) were either known to be associated with cardiac disease, or predicted to be damaging by PolyPhen-2 [34], (Supp. Table 15), hinting at the importance of an Arg at those positions.

It should be noted that we sometimes found more than one reported mutation at the same residue, *e.g.* R31H, R31C and R31G in MYL3 (Uniprot P08590). Estimation of pathogenicity by Polyphen-2 gave in some cases different predictions, *i.e.* probably damaging for R31H and R31C, while benign for R31G. Our conservative interpretation of these cases was that the Arg was not essential at that particular position, and we only considered the benign mutation, when building Supp. Table 15.

### 3.5 ArgMe co-localises with other PTMs in many cardiac proteins

In an attempt to identify a general mechanism of action of ArgMe in cardiac tissue, we searched our dataset of cardiac proteins undergoing ArgMe for co-occurrence of ArgMe and other PTMs. We searched PTMs databases (Uniprot, Phosphositeplus, [31]) for the presence of PTMs on residues neighbouring ArgMe sites. We found a total of 63 PTMs in residues that resided within  $\pm 5$  positions of 41 out of the 109 ArgMe sites considered (103 new plus 11 previously known and excluding the 5 ArgMe sites in titin). These PTMs included phosphorylation (Ser, Thr, or Tyr), Lys modifications (ubiquitination, acetylation, methylation), and ArgMe (Supp. Table 15).

### 3.6 Preliminary observations suggest significant changes in protein ArgMe in a cell model of hypertrophic cardiomyopathy

We noted that many of the genetic mutations associated with disease in Supp. Table 15 had been linked to hypertrophic cardiomyopathy. To provide experimental proof-of-concept that ArgMe can be relevant to cardiac disease, we used a recently established model of cardiac-like cell hypertrophy [28, 29]. We treated H9c2 cells with glucose (30 mM for 72 h), and analysed cell size and global ArgMe patterns. We observed an increase in H9c2 cell size by 37% (Figure 3A and Supp. Figure 1), with concomitant slight, and general, reduction in protein mono- and diArgMe in glucose-treated H9c2 cells (Figure 3B). We also noticed that the intensity of several protein bands significantly decreased, or bands were absent, after glucose treatment (filled arrows, Figure 3B), which was particularly obvious when visualising protein monoArgMe. In contrast, a protein band at *ca.* 31 KDa seemed to increase after glucose treatment (dotted arrow, Figure 3B).

## 4. Discussion

In this work we have identified proteins undergoing ArgMe in 13 human organs, based on published proteomic data. We propose that ArgMe events are mainly defined at the tissue level, although some ArgMe seems to derive from early embryonic events. In the heart, we found 58 proteins that are

Accepted Article

methyated at 103 sites in both fetal and adult samples. Analysis of these data and our experiments using a cardiac-like cell model of hypertrophy suggest that ArgMe can play important roles in cardiomyocyte physiology.

From a proteomic point of view, the scope of ArgMe in tissues has only recently begun to be explored. The availability of specific antibodies has revolutionised the search for global ArgMe marks beyond the histone code [3, 4, 35]; however, these experiments are very demanding in resources and time. Here, we generated a map of protein ArgMe in the human body by data mining. To our knowledge, the most relevant study to quality control our data is that from Guo *et al* [3], which described the arginine methylome of the mouse brain. In this work, researchers identified a total of 807 monoArgMe sites (453 proteins) and 697 diArgMe sites (321 proteins). This compares with our list of 232 monoArgMe sites (177 proteins) and 139 diArgMe sites (109 proteins) in the human brain, including 60 proteins also identified by Guo *et al*. On the one hand, these numbers suggest that we could be underestimating the total number of ArgMe sites in our analyses. This could indeed be the case as our approach is clearly biased towards abundant proteins, while Guo *et al* enriched their samples in arginine methylated peptides [3]. On the other hand, Guo *et al* used a rather high precursor mass tolerance of 50 ppm for peptide identification using an Orbitrap Elite or a Q-Exactive mass spectrometer, which could lead to high false positive discovery rates, while we used 4.5 ppm. Also, it is conceivable that some ArgMe sites are species-specific (*e.g.* R680 in human Nav1.5 is replaced by H680 in mouse Nav1.5). Finally, our approach depends on the quality of primary proteomic data, for example on the total number of proteins identified. For instance, we identified *ca.* 4900 proteins in the brain, compared to 7000-8000 in *e.g.* pancreas, prostate, retina, and testis. While acknowledging these limitations, our data provide for the first time a framework to picture ArgMe in the human body.

In the original publication by the group of Pandey [24], Desmoplakin was identified by 81 unique peptides in the adult heart, however, no Desmoplakin methylated peptides were detected in our analysis. Moreover, GATA4, Nav1.5 and PGC-1 $\alpha$  were not identified in the original work [24], probably reflecting low protein abundance. Despite these limitations, our results increase the number of proteins thought to be methylated in the human heart from 4 (Desmoplakin, GATA4, Nav1.5, and

PGC-1 $\alpha$ ) to 62. PRMT1, -3, -4, and -5 are the most abundant methyltransferases in fetal and adult heart [24], and must be responsible for most of this ArgMe activity. This is consistent with previous work identifying the PRMTs that methylate Desmoplakin, GATA4, Nav1.5 and PGC-1 $\alpha$  [6, 7, 9, 11].

Bulk ArgMe activity must be defined mainly at the level of the organ, as most (53-58%) ArgMe events were tissue-specific, although this could also reflect tissue-specific protein expression in some cases. We observed that 7-13% ArgMe sites were conserved in most tissues within a particular germ layer. As expected, many proteins undergoing these core ArgMe events (Table 2) were involved in RNA binding and processing [25], and we also noted the presence of proteins (and enrichment in GO terms) related to glycolysis. We propose that core ArgMe events are programmed from early embryonic stages, and that the lack of ArgMe at these sites / proteins could be responsible for some of the serious consequences of PRMT deletion as observed in animal models [14-19]. In support of this, loss of hnRNPs and PABPC1 ArgMe is likely to contribute to embryonic death of PRMT1 [15, 36], and PRMT4 [18] null mice, respectively, and indeed we found that hnRNPs and PABPC1 undergo several core ArgMe events in normal human tissues (Table 2).

Bearing the above in mind, we hypothesised that ArgMe sites conserved along fetal and adult tissues must be particularly relevant to tissue function. Using the heart as a model to test this hypothesis, we found that proteins undergoing ArgMe in both fetal and adult cardiac tissue were specifically rich in GO terms essential to cardiomyocyte identity, relative to the common cardiac proteome. This supported a fundamental role of ArgMe in cardiomyocyte physiology, as proteins participating in critical cardiomyocyte functions (including contraction) and structures (*e.g.* Z-disk) seemed to be more likely to undergo ArgMe. In support of this, we found many disease-associated mutations affecting an Arg position shown to be methylated in both fetal and adult cardiac samples. It is tempting to speculate that, at least in some cases, the pathogenicity of the mutation may be due to the lack of protein regulation at that position by ArgMe.

Building up from the observations that a) several proteins undergoing core ArgMe events were related to glucose metabolism (Table 2), and b) many of the mutations at ArgMe sites were associated with hypertrophic cardiomyopathy (Supp. Table 15), we decided to use a model of cardiac

cell hypertrophy induced by high glucose to experimentally link ArgMe to cardiac disease. H9c2 cells show similar hypertrophic responses to primary neonatal cardiomyocytes [37], and are the most widely used *in vitro* model of cardiomyocyte hypertrophy [38]. In our hands, treatment of H9c2 cells with 30 mM glucose for 3 days led to 37% increase in size, which is consistent with previous reports [28, 29]. We observed a general, subtle, decrease in ArgMe in glucose-treated H9c2 cells, which we rationalise by a decrease in PRMT and SAM cellular concentrations caused by the rapid cell growth, and leading to parallel reduction in ArgMe activity. In combination with this, the fact that protein synthesis is enhanced in hypertrophic cells [39] could lead to an apparent “dilution” of previous ArgMe at the total protein level.

On the other hand, we observed specific events as the intensity of several protein bands was remarkably different upon glucose-induced cell hypertrophy. Although the identity of these proteins is currently unknown, our results suggest that specific changes in protein ArgMe occur following induction of cardiac-like cell hypertrophy using glucose. However, given the associations of PRMTs with cellular metabolism (both as regulating and regulated enzymes, [2]) it may be that other pathways contribute to the observed differences in ArgMe, and clearly many more studies are needed to establish a causal relationship between hypertrophy and abnormal ArgMe, or *viceversa*. In this respect, it has recently been shown that knockdown of PRMT5 leads to cardiomyocyte hypertrophy (through mechanisms that include GATA4 ArgMe-Lys acetylation interplay) [7]. Taking together our analysis of genetic mutations at ArgMe sites, and our pilot H9c2 experiments, we propose that abnormal ArgMe of specific proteins could be both cause and consequence of cardiac cell hypertrophy.

ArgMe alters the hydrophobicity, size and bonding of Arg residues. These changes can affect protein-protein interactions and, for instance, recognition of proteins by enzymes catalysing other PTMs. For example, Na<sub>v</sub>1.5 R513 methylation and S516 phosphorylation regulate each other, with possible implications in cardiac disease [40]. Likewise, PGC-1 $\alpha$  ArgMe / acetylation equilibrium is believed to play a role in cardiomyopathy associated to metabolic disorders [41]. Additionally, GATA4 ArgMe inhibits its acetylation [7]; while loss of ArgMe at R2834 reduces Desmoplakin

phosphorylation at S2839 [6]. Based on our findings and the above discussion, we propose that ArgMe cross-talks with other PTMs in at least 1/3 of the cases. However, ArgMe seems to be a stand-alone protein in 2/3 of the methylation sites. This is perhaps not surprising, as ArgMe is a stable PTM [2, 25], which reduces the degrees of freedom for PTM interplays. We predict that systems approaches to ArgMe-PTMs cross-talk will increasingly incorporate engineered PRMTs and SAM analogues for target identification [42], followed by the study of the effect of ArgMe on an individual protein basis.

### **Concluding Remarks**

In this work we have a) mapped ArgMe in 13 human tissues and proposed that most ArgMe activity is controlled at the organ level, while core ArgMe events are programmed from early embryonic stages; b) shown that there is a high frequency of cardiac disease-associated mutations at ArgMe sites conserved in fetal and adult heart, and that GO terms essential to cardiomyocyte ontology are enriched in these proteins; and c) provided experimental evidence that specific proteins undergo significant changes in ArgMe upon induction of cardiac-like cell hypertrophy.

### **Acknowledgements**

DOO acknowledges Rivers state university of Science & Technology Port Harcourt Nigeria, and TETFund Nigeria (Academic staff Training and development unit) for funding. We thank Kath Bulmer and Sage Pickwell for excellent technical assistance. We thank Cristina Chiva (CRG / UPF Proteomics Unit, Barcelona) for useful discussions.

### **Conflict of interest statement**

The authors have declared no conflict of interest.

## 5. References

1. Kensen ON. Interpreting the protein language using proteomics. *Nat Rev Mol Cell Biol.* 2006, 7:391–403
2. Morales Y, Cáceres T, May K, Hevel JM. Biochemistry and regulation of the protein arginine methyltransferases (PRMTs). *Arch Biochem Biophys.* 2016;590:138-52.
3. Geoghegan V, Guo A, Trudgian D, Thomas B, Acuto O. Comprehensive identification of arginine methylation in primary T cells reveals regulatory roles in cell signalling. *Nat Commun.* 2015;6:6758.
4. Guo A, Gu H, Zhou J, Mulhern D, et al. Immunoaffinity enrichment and mass spectrometry analysis of protein methylation. *Mol Cell Proteomics.* 2014;13(1):372-87.
5. Khoury GA, Baliban RC, Floudas CA. Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. *Sci Rep.* 2011, 1.
6. Albrecht LV, Zhang L, Shabanowitz J, Purevjav E, et al. GSK3- and PRMT-1-dependent modifications of desmoplakin control desmoplakin-cytoskeleton dynamics. *J Cell Biol.* 2015;208(5):597-612.
7. Chen M, Yi B, Sun J. Inhibition of cardiomyocyte hypertrophy by protein arginine methyltransferase 5. *J Biol Chem.* 2014;289(35):24325-35.
8. Beltran-Alvarez P, Pagans S, Brugada R. The cardiac sodium channel is post-translationally modified by arginine methylation. *J Proteome Res.* 2011;10(8):3712-9.
9. Beltran-Alvarez P, Espejo A, Schmauder R, Beltran C, et al. Protein arginine methyltransferases-3 and -5 increase cell surface expression of cardiac sodium channel. *FEBS Lett.* 2013;587(19):3159-65.
10. Beltran-Alvarez P, Tarradas A, Chiva C, Pérez-Serra A, et al. Identification of N-terminal protein acetylation and arginine methylation of the voltage-gated sodium channel in end-stage heart failure human heart. *J Mol Cell Cardiol.* 2014;76:126-9.

11. Teyssier C, Ma H, Emter R, Kralli A, Stallcup MR. Activation of nuclear receptor coactivator PGC-1alpha by arginine methylation. *Genes Dev.* 2005;19(12):1466-73.
12. Pasternack DA, Sayegh J, Clarke S, Read LK. Evolutionarily divergent type II protein arginine methyltransferase in *Trypanosoma brucei*. *Eukaryot Cell.* 2007, 6(9):1665-81.
13. Krause CD, Yang ZH, Kim YS, Lee JH, et al. Protein arginine methyltransferases: evolution and assessment of their pharmacological and therapeutic potential. *Pharmacol Ther.* 2007;113(1):50-87.
14. Yu Z, Chen T, Hébert J, Li E, Richard S. A mouse PRMT1 null allele defines an essential role for arginine methylation in genome maintenance and cell proliferation. *Mol Cell Biol.* 2009 Jun;29(11):2982-96.
15. Pawlak MR, Scherer CA, Chen J, Roshon MJ, Ruley HE. Arginine N-methyltransferase 1 is required for early postimplantation mouse development, but cells deficient in the enzyme are viable. *Mol Cell Biol.* 2000;20(13):4859-69.
16. Tsai YJ, Pan H, Hung CM, Hou PT, et al. The predominant protein arginine methyltransferase PRMT1 is critical for zebrafish convergence and extension during gastrulation. *FEBS J.* 2011;278(6):905-17.
17. Swiercz R, Cheng D, Kim D, Bedford MT. Ribosomal protein rpS2 is hypomethylated in PRMT3-deficient mice. *J Biol Chem.* 2008;282(23):16917-23.
18. Yadav N, Lee J, Kim J, Shen J, et al. Specific protein methylation defects and gene expression perturbations in coactivator-associated arginine methyltransferase 1-deficient mice. *Proc Natl Acad Sci USA.* 2003;100(11):6464-8.
19. Tee WW, Pardo M, Theunissen TW, Yu L, et al. Prmt5 is essential for early mouse development and acts in the cytoplasm to maintain ES cell pluripotency. *Genes Dev.* 2010;24(24):2772-7.
20. Iwasaki H, Kovacic JC, Olive M, Beers JK, et al. Disruption of protein arginine N-methyltransferase 2 regulates leptin signaling and produces leanness in vivo through loss of STAT3 methylation. *Circ Res.* 2010;107(8):992-1001.

21. Neault M, Mallette FA, Vogel G, Michaud-Levesque J, Richard S. Ablation of PRMT6 reveals a role as a negative transcriptional regulator of the p53 tumor suppressor. *Nucleic Acids Res.* 2012;40(19):9513-21.
22. Jeong HJ, Lee HJ, Vuong TA, Choi KS, et al. Prmt7 Deficiency Causes Reduced Skeletal Muscle Oxidative Metabolism and Age-Related Obesity. *Diabetes.* 2016;65(7):1868-82.
23. Kim JD, Park KE, Ishida J, Kako K, et al. PRMT8 as a phospholipase regulates Purkinje cell dendritic arborization and motor coordination. *Sci Adv.* 2015;1(11):e1500615.
24. Kim MS, Pinto SM, Getnet D, Nirujogi RS, et al. A draft map of the human proteome. *Nature.* 2014;509(7502):575-81.
25. Bedford MT, Clarke SG. Protein arginine methylation in mammals: who, what, and why. *Mol. Cell* 2009;33(1):1–13.
26. Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics.* 2009;10:48.
27. Eden E, Lipson D, Yogev S, Yakhini Z. Discovering motifs in ranked lists of DNA sequences. *PLoS Comput Biol.* 2007;3(3):e39.
28. Liang D, Zhong P, Hu J, Lin F, et al. EGFR inhibition protects cardiac damage and remodeling through attenuating oxidative stress in STZ-induced diabetic mouse model. *J Mol Cell Cardiol.* 2015;82:63-74.
29. Zhong P, Wu L, Qian Y, Fang Q, et al. Blockage of ROS and NF- $\kappa$ B-mediated inflammation by a new chalcone L6H9 protects cardiomyocytes from hyperglycemia-induced injuries. *Biochim Biophys Acta.* 2015;1852(7):1230-41.
30. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis". *Nature Methods.* 2012;9:671-675.
31. Hornbeck PV, Zhang B, Murray B, Kornhauser JM, et al. PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res.* 2015;43(Database issue):D512-20.

32. Stenson PD, Mort M, Ball EV, Shaw K, et al. The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. *Hum Genet.* 2014;133(1):1-9.
33. Campuzano O, Sanchez-Molero O, Mademont-Soler I, Riuró H, et al. Rare Titin (TTN) Variants in Diseases Associated with Sudden Cardiac Death. *Int J Mol Sci.* 2015;16(10):25773-87.
34. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, et al. A method and server for predicting damaging missense mutations. *Nat Methods.* 2010;7(4):248-9.
35. Gu H, Ren JM, Jia X, Levy T, et al. Quantitative Profiling of Post-translational Modifications by Immunoaffinity Enrichment and LC-MS/MS in Cancer Serum without Immunodepletion. *Mol Cell Proteomics.* 2016;15(2):692-702.
36. Gurunathan G, Yu Z, Coulombe Y, Masson JY, Richard S. Arginine methylation of hnRNPUL1 regulates interaction with NBS1 and recruitment to sites of DNA damage. *Sci Rep.* 2015;5:10475.
37. Watkins SJ, Borthwick GM, Arthur HM. The H9C2 cell line and primary neonatal cardiomyocyte cells show similar hypertrophic responses in vitro. *In Vitro Cell Dev Biol Anim.* 2011;47(2):125-31.
38. Valdés-Mas R, Gutiérrez-Fernández A, Gómez J, Coto E, et al. Mutations in filamin C cause a new form of familial hypertrophic cardiomyopathy. *Nat Commun.* 2014;29(5):5326.
39. Saeedi R, Saran VV, Wu SS, Kume ES, et al. AMP-activated protein kinase influences metabolic remodeling in H9c2 cells hypertrophied by arginine vasopressin. *Am J Physiol Heart Circ Physiol.* 2009;296(6):H1822-32.
40. Beltran-Alvarez P, Feixas F, Osuna S, Díaz-Hernández R, et al. Interplay between R513 methylation and S516 phosphorylation of the cardiac voltage-gated sodium channel. *Amino Acids.* 2015;47(2):429-34.

41. Garcia MM, Guéant-Rodriguez RM, Pooya S, Brachet P, et al. Methyl donor deficiency induces cardiomyopathy through altered methylation/acetylation of PGC-1 $\alpha$  by PRMT1 and SIRT1. *J Pathol.* 2011;225(3):324-35.
42. Guo H, Wang R, Zheng W, Chen Y, et al. Profiling substrates of protein arginine N-methyltransferase 3 with S-adenosyl-L-methionine analogues. *ACS Chem Biol.* 2014;9(2):476-84.

**Supplementary Tables 1-13.** For each tissue, in alphabetical order, an excel file shows in different sheets total proteins identified, monoArgMe sites and diArgMe sites as retrieved from MaxQuant analyses.

**Supplementary Table 14.** For fetal heart, an excel file shows in different sheets total proteins identified, monoArgMe sites and diArgMe sites as retrieved from MaxQuant analyses.

**Supplementary Table 15.** List of non-histone proteins and ArgMe sites conserved in both fetal and in adult heart, mutations at ArgMe sites, and co-occurrence of other PTMs. Uniprot codes and gene names are shown, sorted by Uniprot code. Desmoplakin (P15924, DSP), GATA4 (P43694), Nav<sub>v</sub>1.5 (Q14524, SCN5A) and PGC-1 $\alpha$  (Q9UBK2, PPARGC1A) are included. The star (\*) symbol by the Uniprot code represents that the methylation site(s) has previously been reported in either the Uniprot or the Phosphositeplus Database [31]. PTMs are shown as follows: S-p, T-p or Y-p = phosphorylated Ser, Thr or Tyr; K-ac = acetylated Lys; K-m1 or m2 = mono- or dimethylated Lys; K-ub = ubiquitinated Lys; R-m1 or m2 = mono- or dimethylated Arg.

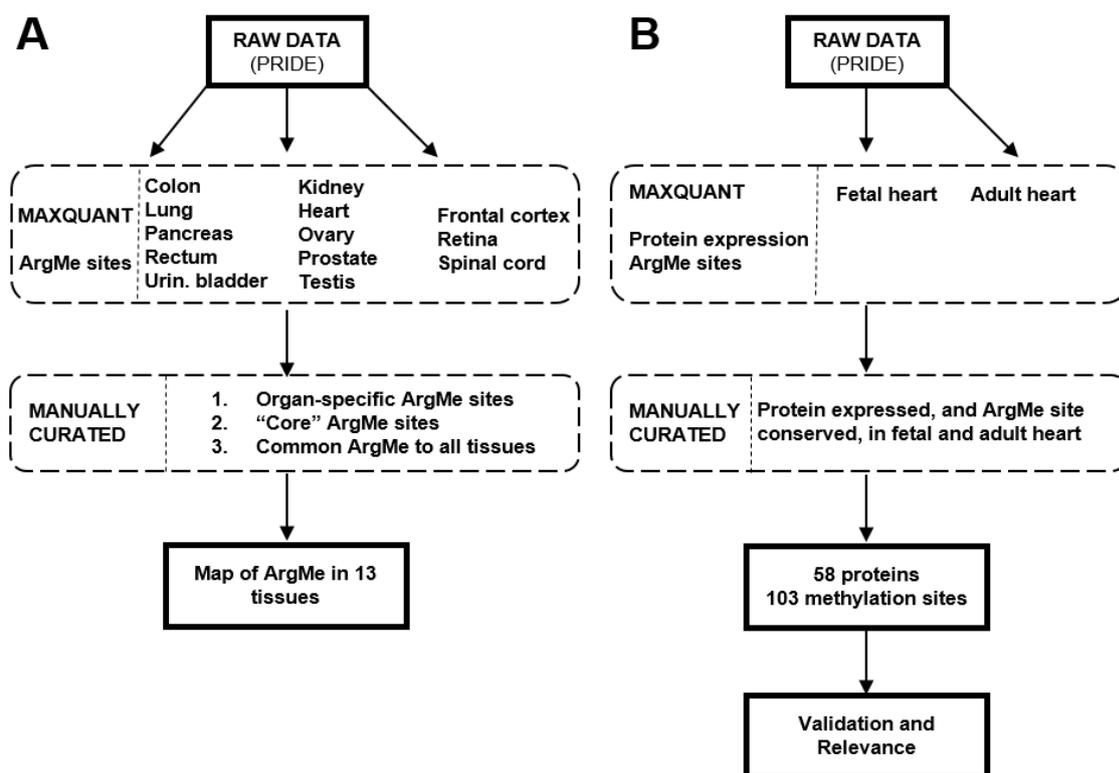
**Supplementary Table 16.** GO term enrichment analysis of proteins in Supplementary Table 15. Biological Process, Molecular Function, and Component ontologies are shown in separate sheets.

**Supplementary Figure 1.** **A.** Images of control H9c2 cells used to quantify cell surface. **B.** Images of glucose-treated H9c2 cells used to quantify cell surface.

## Legends

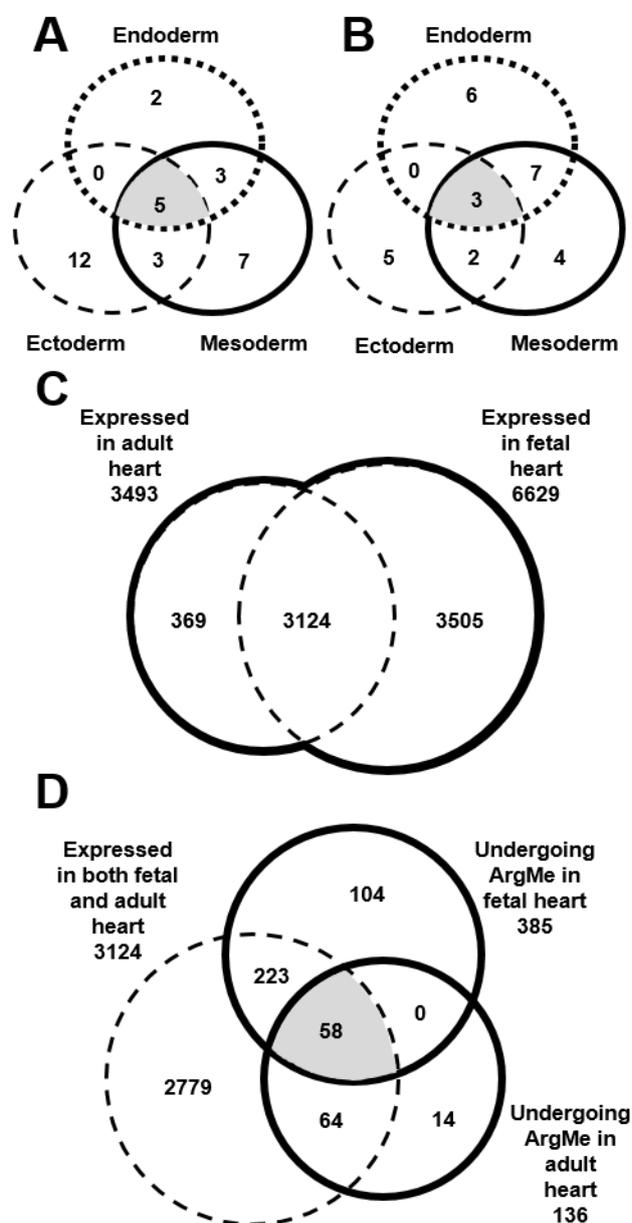
**Figure 1. A.** Flow chart illustrating our protocol to identify ArgMe sites in proteins from 13 human tissues. “Core” ArgMe sites were defined as those common to at least 4 tissues in endoderm- or mesoderm-derived tissues, or conserved throughout the 3 ectoderm-derived tissues. **B.** Flow chart illustrating our protocol to identify ArgMe sites conserved in fetal and adult heart.

**Figure 1**



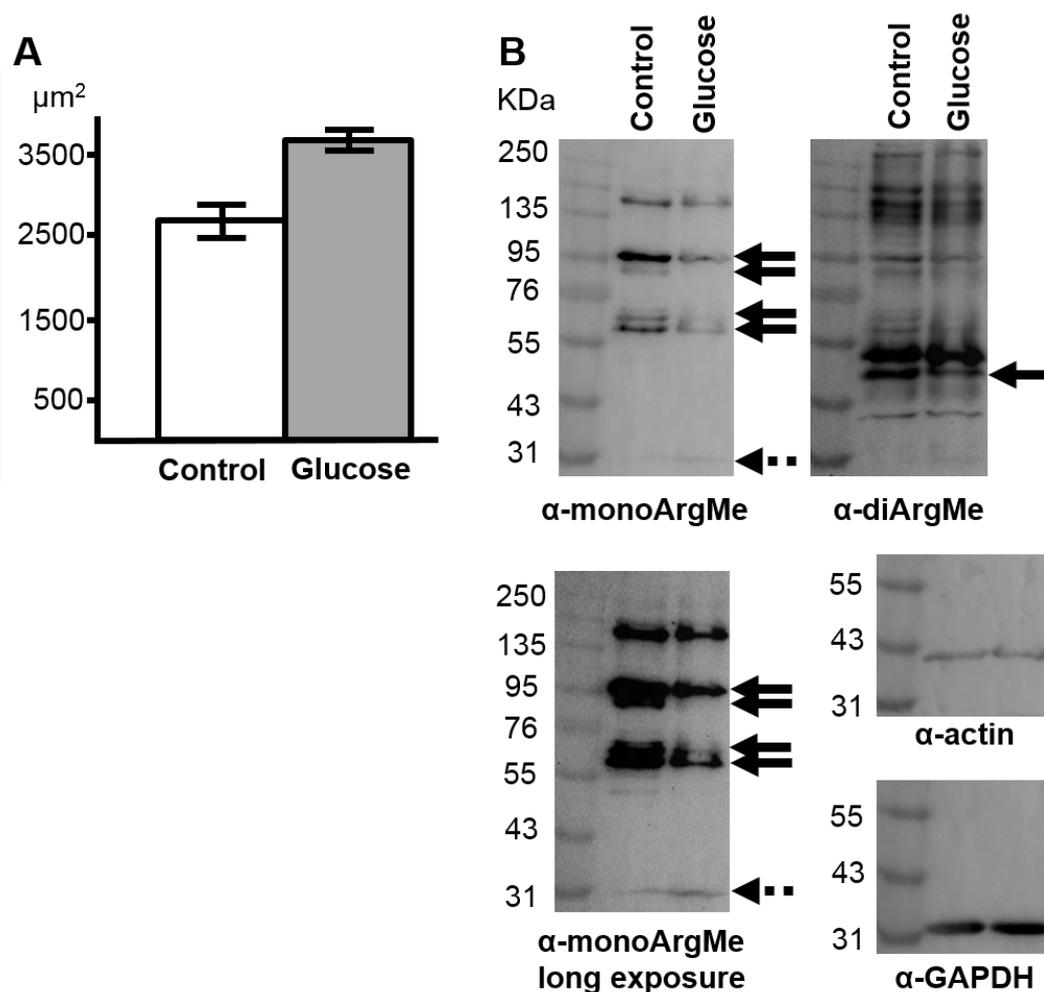
**Figure 2. A.** Venn-diagram comparing core monoArgMe sites conserved in tissues derived from the 3 germ layers. **B.** Venn-diagram comparing core diArgMe sites conserved in tissues derived from the 3 germ layers **C.** Venn-diagram comparing the total number of proteins identified as expressed in adult and fetal heart. **D.** Identification of 58 proteins undergoing ArgMe in adult and fetal heart (shaded intersection). 103 methylation sites in these 58 proteins were conserved in adult and fetal samples.

**Figure 2**



**Figure 3. A.** Chart showing significant ( $P < 0.05$ ) increase in cell size upon induction of H9c2 hypertrophy by glucose. Control:  $2740 \pm 202 \mu\text{m}^2$  ( $n = 94$  cells), Glucose-treated:  $3743 \pm 102 \mu\text{m}^2$  ( $n = 100$  cells), **B.** ArgMe profiles of H9c2 cells (Control lane) and upon glucose-induced hypertrophy (Glucose lane). The membrane was sequentially blotted with antibodies targeting mono- and diArgMe, as well as actin and GAPDH as protein loading controls. Specific proteins (filled arrows) underwent remarkable reduction in ArgMe, and a protein at 31 KDa (dotted arrow) showed increased monoArgMe, after glucose treatment.

**Figure 3**



**Table 1.** Number of mono- and diArgMe sites in tissues classified as per germ layer origin. For each tissue, the total number of mono and diArgMe sites is shown. The number of core ArgMe sites as well as the % this represents of the total number of mono- or diArgMe sites for that tissue are shown between brackets.

**Table 1**

Germ layer origin	Tissue	No. Mono-ArgMe sites (No., % of core sites within germ layer)	No. Di-ArgMe sites (No., % of core sites within germ layer)
<b>Endo</b>	Colon	151 (10, 7%)	129 (19, 12%)
	Lung	103 (10, 10%)	75 (19, 21%)
	Pancreas	153 (10, 7%)	173 (19, 9%)
	Rectum	157 (10, 6%)	126 (19, 12%)
	Ur. bladder	195 (10, 5%)	144 (19, 11%)
<b>Meso</b>	Kidney	279 (18, 6%)	74 (21, 22%)
	Heart	221 (18, 8%)	61 (21, 26%)
	Ovary	297 (18, 6%)	286 (21, 6%)
	Prostate	279 (18, 6%)	128 (21, 12%)
	Testis	205 (18, 9%)	146 (21, 11%)
<b>Ecto</b>	Fr. cortex	234 (20, 8%)	140 (13, 7%)
	Retina	462 (20, 4%)	166 (13, 6%)
	Spinal cord	232 (20, 9%)	54 (13, 19%)
	<b>Average</b>	228 (15, 7%)	131 (15, 13%)

**Table 2.** Proteins undergoing core ArgMe sites in each germ layer, or combination of them, sorted alphabetically by gene name for each section. Type of ArgMe refers to monoArgMe (m1) or diArgMe (m2). Novelty of the ArgMe was assessed using Phosphositeplus as a reference [31].

Table 2

Germ layer	ENSEMBL ID	Gene Name	Gene Description	ArgMe site, type	Novel type	
Endo + Meso + Ecto	ENSP00000421592	ALYREF	Aly/REF export factor	38, m2	No	
	ENSP00000295006	CAPN2	Calpain 2, (m/II) large subunit	23, m1	Yes	
	ENSP00000339063	EEF1A1	Eukaryotic translation elongation factor 1 $\alpha$ 1	166, m1, m2	No	
	ENSP00000408907	HSPA1A	Heat shock protein family A (Hsp70) 1A	469, m1	No	
Endo + Meso	ENSP00000377440	RAB5C	RAB5C, member RAS oncogene family	4, m1	Yes	
	ENSP00000409681	C11orf68	UPF0696 protein c11orf68	14, m2	Yes	
	ENSP00000359717	FHL1	Four and a half LIM domains 1	5, m2	Yes	
	ENSP00000369594	FUS	FUS RNA binding protein	215, 217, m2	No	
	ENSP00000354021	HNRNPA2 B1	Heterogeneous nuclear ribonucleoprotein A2/B1	213, m1	No	
	ENSP00000380446	PABPN1	Poly(A) binding protein, nuclear 1	17, m1, m2	No	
	ENSP00000376899	PTGFRN	Prostaglandin F2 receptor inhibitor	280, m1	Yes	
	ENSP00000365950	RBM3	RNA binding motif (RNP1, RRM) protein 3	105, m2	No	
	ENSP00000349748	SFPQ	Splicing factor proline/glutamine-rich	19, 25, m2	No	
	ENSP00000417952	TFG	TRK-fused gene protein	385, m2	Yes	
Meso + Ecto	ENSP00000455917	ALDOA	Aldolase, fructose-bisphosphate A	56, m1	Yes	
	ENSP00000380070	GAPDH	Glyceraldehyde-3-phosphate dehydrogenase	248, m1	Yes	
	ENSP00000283179	HNRNPU	Heterogeneous nuclear ribonucleoprotein U	733, 739, m2	No	
	ENSP00000313007	PABPC1	Poly(A) binding protein 1	493, m2	No	
Endo	ENSP00000398599	YWHAZ	14-3-3 protein, zeta	41, m1	Yes	
	ENSP00000331699	EWSR1	EWS RNA binding protein 1	563, m2	No	
	ENSP00000301785	HNRNPUL 2	Heterogeneous nuclear ribonucleoprotein U-like 2	656, 684, m2	No	
	ENSP00000349748	SFPQ	Splicing factor proline/glutamine-rich	693, m1	No	
	ENSP00000435614	SLC22A25	Solute carrier family 22 member 25	84, m2	Yes	
	ENSP00000484236	SLC38A5	Solute carrier family 38, member 5	349, m2	Yes	
	ENSP00000400591	SNRPE	Small nuclear ribonucleoprotein protein E	4, m2	Yes	
	ENSP00000432282	TAGLN	Transgelin	183, m1	No	
	ENSP00000465435	ZMYND15	Zinc finger, MYND-type containing 15	145, m2	Yes	
	Meso	ENSP00000451979	APEX1	APEX nuclease	156, m1	Yes
		ENSP00000455329	CRIP1	Cysteine-rich protein 1	68, m1	No
		ENSP00000484789	HIST1H4E	Histone cluster 1, h4e	55, m1	No
ENSP00000354021		HNRNPA2 B1	Heterogeneous nuclear ribonucleoprotein A2/B1	213, m2	No	
ENSP00000313327		HNRNPD	Heterogeneous nuclear ribonucleoprotein D	296, m2	Yes	
ENSP00000215909		LGALS1	Lectin, galactoside-binding, soluble, 1	19, m1	Yes	
ENSP00000465404		PARK7	Parkinson protein 7	48, m1	Yes	
ENSP00000270142		SOD1	Superoxide dismutase 1	70, 116, m1	Yes	
ENSP00000258962		SRSF1	Serine/arginine-rich splicing factor 1	97, m1; 109, m1, m2	No	
Ecto		ENSP00000216254	ACO2	Aconitase 2	84, m1	Yes
	ENSP00000378731	ALDOA	Aldolase, fructose-bisphosphate A	134, m1	Yes	
	ENSP00000418795	CFAP44	Cilia- and flagella-associated protein 44	374, m1	Yes	
	ENSP00000309539	DPYSL2	Dihydropyrimidinase like 2	227, 238, m1	Yes	
	ENSP00000258198	DYNC1L2	Dynein, cytoplasmic 1, light chain 2	397, m1	No	
	ENSP00000379069	G3BP2	Gtpase activating protein (SH3 domain) binding protein 2	468, m2	No	
	ENSP00000253408	GFAP	Glial fibrillary acidic protein	300, m1	Yes	
	ENSP00000353552	HNRNPK	Heterogeneous nuclear ribonucleoprotein K	299, 303, m2	No	
	ENSP00000281938	HSPB8	Heat shock protein family B member 8	78, m2	No	
	ENSP00000420813	PCMT1	Protein-L-isoaspartate O-methyltransferase	182, m1	Yes	
	ENSP00000362413	PGK1	Phosphoglycerate kinase 1	206, m1	Yes	
	ENSP00000371388	SRXN1	Sulfiredoxin 1	16, m2	No	
	ENSP00000379933	TPI1	Triosephosphate isomerase 1	53, m1	Yes	
	ENSP00000336799	TUBA1B	Tubulin alpha 1b	221, m1	Yes	
	ENSP00000446007	VIM	Vimentin	304, m1	Yes	
	ENSP00000462595	WSB1	WD repeat and SOCS box-containing 1	7, m2	Yes	
	ENSP00000487356	YWHAE	14-3-3 protein, epsilon	4, m1	Yes	