Appendix C

details of statistical analysis, protein-protein interaction analysis and commentaries to CDC5L and DAAM2 proteins

PERINATAL CHANGES IN PLASMA OF TYPE 2 DIABETES MELLITUS PREGNANCIES WHO DELIVERED NEWBORNS WITH CARDIOMYOPATHY, DEPRESSION OF THE CENTRAL NERVOUS SYSTEM, OR HEPATOMEGALY

Appendix C

Preliminary statistical analysis

Taking into account that in the vast majority, newborns have a combination of signs of fetopathy, initially each group of patients was divided into the corresponding number of subgroups, the number of which was 12 (excluding the control group). However, the size of such subgroups is too small and averages 7.13 people, which is a serious obstacle for the further. In this regard, either a reduction in subgroups is required, or patients are only counted for one, the main, sign of fetopathy. The reduction of subgroups is not possible in this case, since initially we have a fixed number of differentiated types of fetopathy. At the same time, patient accounting for only one type of fetopathy requires a preliminary analysis of the total proteome, which would determine the most pronounced type of diabetic fetopathy in a given newborn, and, thus, would allow regrouping within the studied cohort of patients.

To carry out the regrouping of patients according to the main most manifested) type of fetopathy and compaction of stratification, 30 proteins were randomly selected, which are common for each individual studied. Proteins were ranked according to their representation on the basis of the NSFA and cluster analysis was carried out using the statistical method of Spearman rank correlation with determination of the distance between clusters using the “average distance” method. Statistical analysis was performed at a confidence level of $p<0.05$. **Figure 1S** shows the results of a cluster analysis of the studied patients. As highlighted, in addition to the control group (**CNT**), clusters of patients with the most manifested signs of cardiomyopathy (**CRD**), hepatomegaly (**HPM**), and syndrome of CNS depression (**CNS**) were also identified. Unfortunately, a pronounced clustering of the group with a sign of macrosomia, as the main sign of fetopathy, was not obtained. This symptom is distributed only together with other types of fetopathy and was regarded as a concomitant symptom of diabetic fetopathy, which transfers it as a non-specific signs of diabetic fetopathy, the determination of which is difficult or impossible in the context of the available proteomic data. Thus, when challenging and analyzing proteomic data, we will be guided by the stratification of patients according to the main most pronounced signs of fetopathy obtained as a result of cluster analysis, i.e., groups with **CNS ($p = 0.0003$)**, **HPM ($p = 0.0017$)**, **CRD ($p = 0.0008$)** and the control group **CNT ($p = 0.0001$)**. A group with signs of only macrosomia will be ruled out since it does not have significant differences from other signs of diabetic fetopathy **(MCR, $p = 0.1067$)** and appears to be only as a concomitant feature of diabetic fetopathy.

For further consideration, we used the entire set of proteins (complete proteome), as the general set of proteins identified in all patients of the studied cohort. To exclude insignificant proteins and those that could be related to unexplained (or external) factors of the proteomic composition, we used a frequency factor $F>0.8$, which means that the certain protein should be present in at least 80% of the certain studied group and in 80% of patients of the whole study cohort, otherwise, the proteins was not considered as a potential marker or protein with potential diagnostic or prognostic significance. The frequency factor was used due to small size of the studied cohort which carries a risk of implausible reflection of proteins statistical distribution. This restriction was introduced to determine the results obtained only as research.

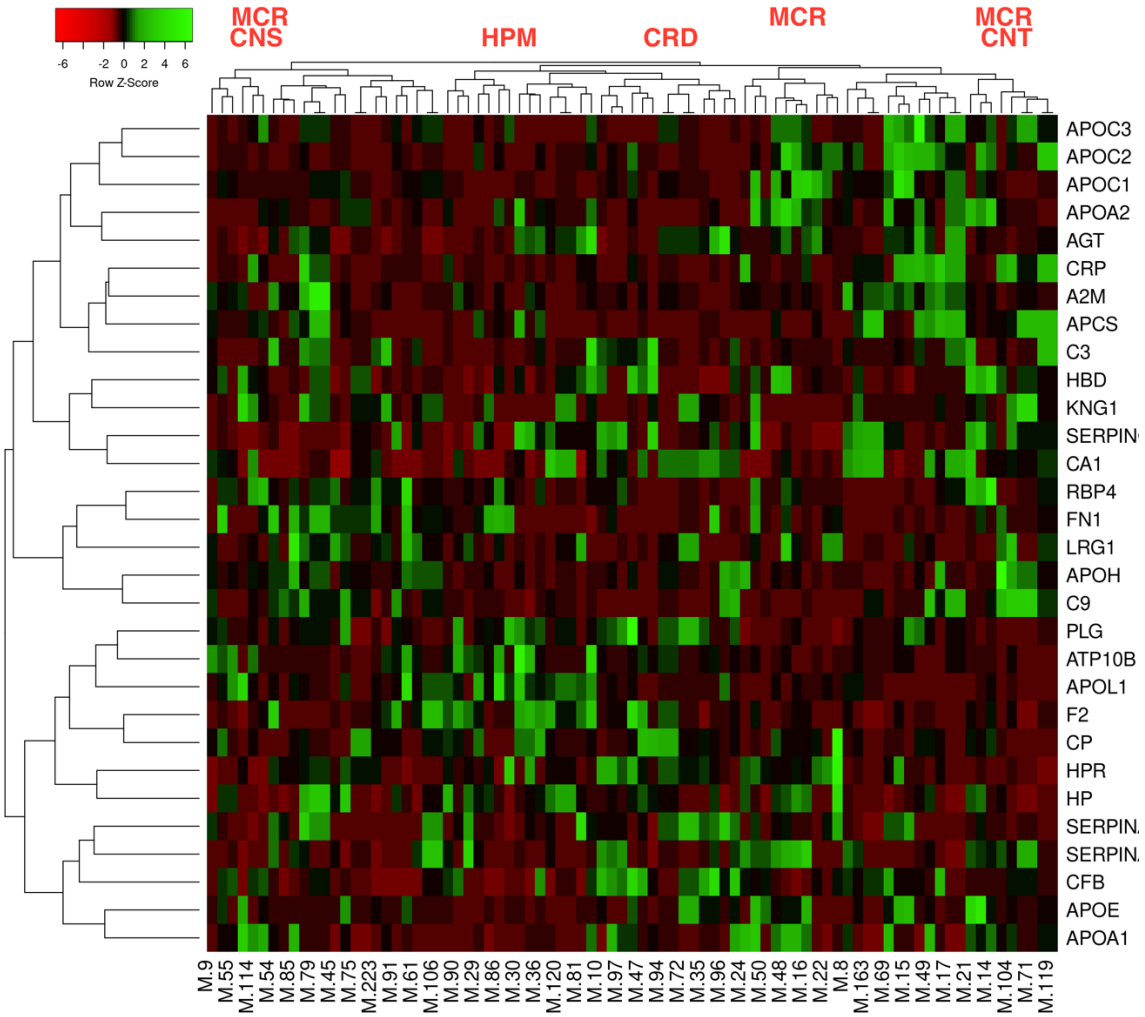

Figure S1: heat map of the cluster analysis of the studied patients. The analysis was provided using randomly selected 30 proteins common to all patients. Cluster analysis was performed using the statistical method of Spearman's rank correlation analysis and determining the distance between patients using the “average distance” method. Cluster analysis divided patients into groups according to the signs of fetopathy CRD, CNS and HPM, and the control group. The sign of macrosomia (MCR) is not segregated, since it is statistically interspersed with other signs of fetopathy, so this symptom cannot be considered as a group-specific. Cluster analysis was performed with a confidence level of $p < 0.05$; for the sign of macrosomia MCR, the level of statistical reliability was $p=0.1067$.

As a result of the analysis, we found 40 biological markers, which can be divided into two groups: adjacent (generalized for any two types of fetopathy) and specific (characterize a certain one type of fetopathy). Moreover, markers generalized for all three types of fetopathy that are studied in this work were not found. Following PCA analysis and analysis of quantitative difference significance, the total number of markers was truncated to 30 proteins. For the CNS group 14 markers were detected; for cardiomyopathy there were 11 specific markers; a minimal group of markers was found for hepatomegaly group (only 4 markers). Among adjustment markers, 3 were shared between CNS and CRD groups;; 4 markers were common for CNS and HPM; and only 2 common markers were found between HPM and CRD (**Table 2 in the main text**). This indicates a great difference in the biological processes that underlie the development of a particular pathology.

APPENDIX-C

DETAILS OF STATISTICAL ANALYSIS, PROTEIN-PROTEIN INTERACTION ANALYSIS AND COMMENTARIRES TO CDC5L AND DAAM2

Protein-protein interactions (PPI)

There are 9 common (adjacent) markers between three groups of diabetic fetopathy: **APOC4** (*Apolipoprotein C-IV*), **LG3BP** (*Galectin-3-binding protein*), **PEDF** (*Pigment epithelium-derived factor*), **TSP1** (*Thrombospondin-1*), **PHLD** (*Phosphatidylinositol-glycan-specific phospholipase D*), **ALS** (Insulin-like growth factor-binding protein complex acid labile subunit), **DPF3** (*Zinc finger protein DPF3*), **DAAM2** (*Disheveled-associated activator of morphogenesis 2*) and **LAMB4** (*Laminin subunit beta-4*). PPI analysis showed that these proteins produce a poorly dense core of functional interactions (the coefficient of protein interactions is *PPI = 0.389*), except for a pair of **PEDF** and **TSP1** and cluster of **DAAM2, DFP3** and **LAMB4**. Assumingly, it indicates involvement of these proteins in the general processes leading to the development and progression of a particular fetopathy. Given the large number of groups of specific markers, most likely, adjacent markers are a reflection of secondary signs of fetopathy, and encompass peripheral biological process. The multiple molecular functions possessed by adjacent markers also support this proposition; these proteins are characterized by pleiotropic property in the cell's life cycle. It is worth noting that one of the most important proteins is **LAMB4**, since this protein is a direct participant in the organization of tissues and the process of organogenesis in the embryonic period and seems critical for determining cell migration and proper orientation in the processes of organogenesis and tissue formation, also fixing cells in the intercellular matrix. Similar functions are performed by **TSP1**, which is paired with **PEDF1**, but its role is largely limited to the adhesive functions that determine the intercellular interaction and the interaction of cells with the extracellular space.

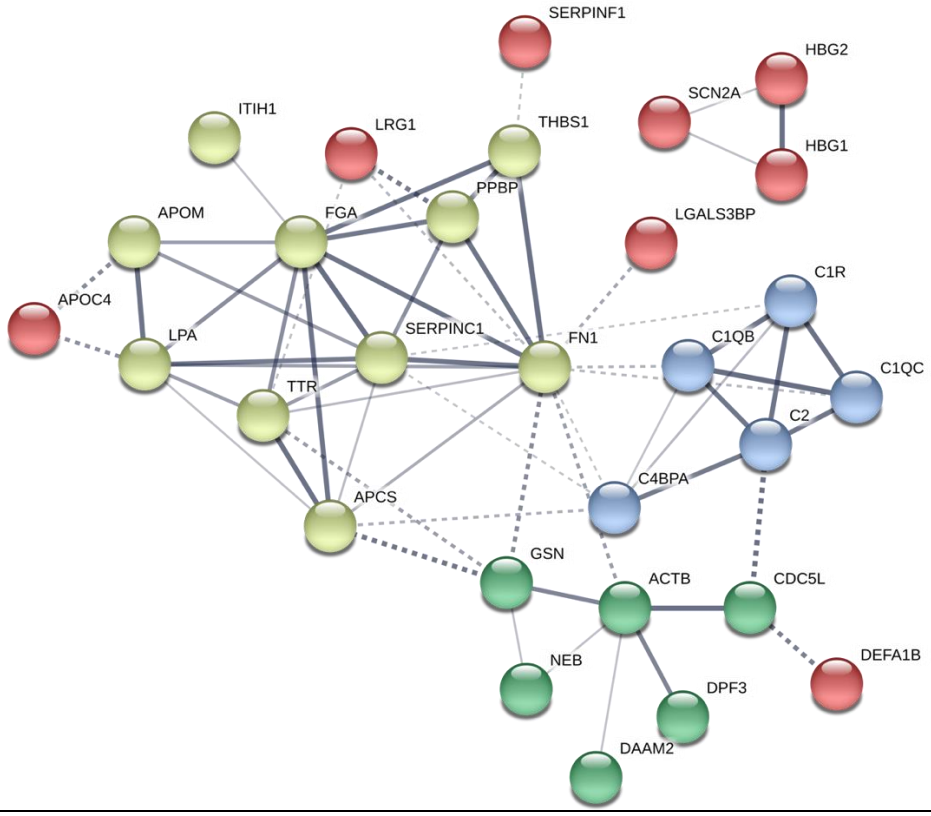

Figure S2: the rout map of the network interaction of proteins included in the groups of individual and adjacent markers of the studied types of fetopathy. Proteins form a compact dense core of network interaction (**PPI = 8.02e-11**). The network can be dividing it into three functional clusters connected with each other by common and transient biological processes. The main cluster is shown **in red** and is determined by the proteins involved in the transmission of signaling molecules, catalytic reactions of the interaction of ligands with receptors, as well as modulation of signaling pathways. Internal network communication is based on the signal transmission and humoral regulation of the immune response, in particular, an autoimmune reaction. Most of the proteins shown in the figure belong to secreted proteins or their catalytic activity is mainly expressed in the extracellular space.

APPENDIX-C

DETAILS OF STATISTICAL ANALYSIS, PROTEIN-PROTEIN INTERACTION ANALYSIS AND COMMENTARIRES TO CDC5L AND DAAM2

Three functional clusters, as shown in Figure S2, are formed by proteins involved in the immune response, proteins involved in the structuring of the intercellular space, and regulatory signaling proteins. Regulatory function is accomplished through the expression of proteolytic catalytic activity, as well as through the transmission of signal (ligands) to receptors along signaling pathways, thereby forming a network of humoral regulation of the immune response and the body's protective response (**GO: 0006952, p = 0.00025**). The vast majority of the proteins participating in the interactions (17 proteins) are secreted proteins, or proteins whose main activity is manifested in the extracellular space region (**GO: 0044421, p = 7.90e-09**), which indicates their high para- and endocrine potentials.

Commentary to CDC5L protein

Murine models demonstrated that CDC5L-/- embryos die rather early and reach only 1.5 days of the development: the nucleus in the cells of such embryos is fragmented, which reflects direct apoptosis processes, which is also expressed in the re-localization of CDC5L from the cell nucleus to the cytoplasm. Heterozygous mice (**Figure S3**) by PLRG1 or CDC5L show a survival rate of not more than 25% on the third day of embryonic development. Using tunneling microscopy for investigation of the surviving embryos it has been shown a dramatic increase in degree of the degraded neurons, especially in the hippocampus, as well as cardiomyocytes, in comparison with the control group of animals.

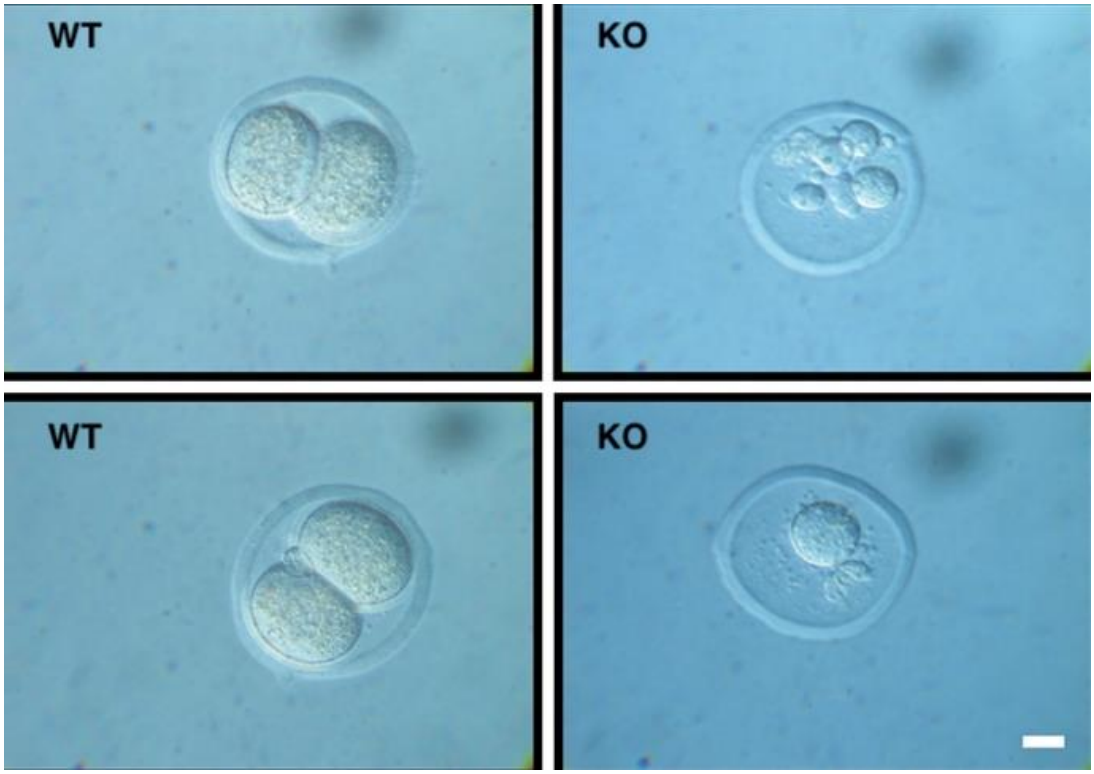

Figure 3S: **this is a reproduced and adapted image from the article [SR1] (SEE FOOTNOTE)**. The two-cell stage of embryonic development (1.5 days) of the normal (wild, heterozygous PLGR1Δ/+) type is shown on the left. On the right is the two-cell stage of

[SR1] This is adapted and reproduced Figure 2 on of the manuscript by Kleinridders A, Pogoda HM, Irlenbusch S, et al. PLRG1 is an essential regulator of cell proliferation and apoptosis during

the embryo with a deficient (with deletion) gene PLRG1Δ/Δ of the embryo, where there is no complexation with CDC5L for DNA repair and the pre-mRNA splicing mechanism is inactive. The cell nucleus is fragmented, cell division is impaired, and embryo are degenerated. The scale (white meniscus in the lower right sector of the figure) is 10 μm.

Commentary to DAAM2 protein

Destruction in molecular functioning or expression of **DAAM2** leads to such consequences as non-compact cardiomyopathy. At least on animal models, it has been highlighted that excessive expression or, on the contrary, a knockout of the DAAM1 or **DAAM2** gene turns to disruption of actin nucleation and cell cytoskeleton organization in embryonic fibroblasts. In summary, it leads to dire consequences of morphogenesis of organs excessively *deep trabecular ventricular myocardial cavities and uncompressed left ventricular myocardium*. In particular, morphogenesis abnormalities are expressed in the orientation of *cardiomyocytes*, where in control histological sections they *are oriented vertical to the walls of the cone*, and in sections with impaired expression and function, **DAAM2** and DAAM1 *cells look round and much smaller in size and mixed with other cells* (**Figure 4S**).

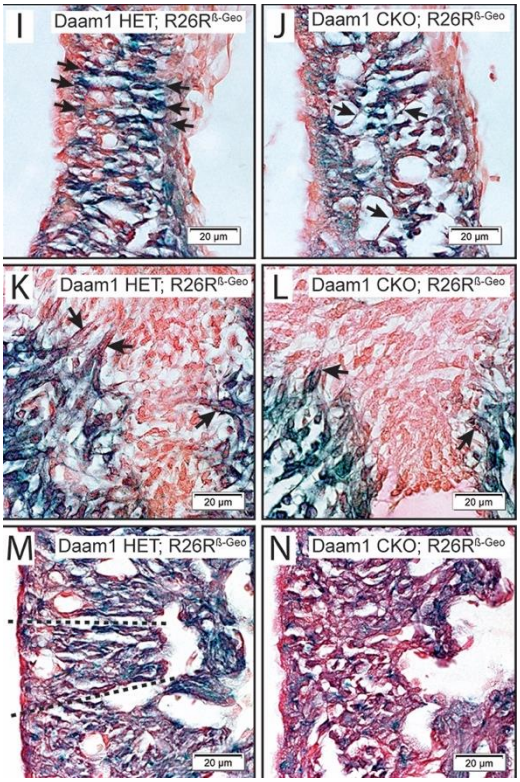

Figure 4S: **the image is reproduced and adapted from [SR2] (SEE FOOTNOTE)**. DAAM1 and DAAM2 are required for ventricular compaction and polarization of cardiomyocytes during embryogenesis. The figure shows stained cardiomyocytes of the control (left panel) samples and samples with impaired functioning of DAAM1 / DAAM2 (right panel; pink color corresponds to eosin staining). In control samples, cardiomyocytes are oriented vertically along the axis of the conotruncal region of the myocardium and have a tight intercellular connection. In experimental samples with impaired expression of DAAM1 / DAAM2, cardiomyocytes are disorganized, mixed with other cells, have a weak intercellular connection, and are not organized along the axis.

Also, on histological stained sections of control samples, demonstrated that cardiomyocytes are firmly connected to each other, while in experimental samples the intercellular interaction is very weak, which is due to inhibition of the synthesis of type-II collagen and its secretion to the extracellular space and a decrease in the polymerization and polarization of actin filaments of the cytoskeleton (Figure 4S). At the same time, an interesting morphological feature is revealed: DAAM1 dysfunction leads to disorganized morphogenesis and functioning, as a rule, of the right ventricle of the heart, while the left ventricle remains, most often, unaffected and morphologically functional.

vertebrate development and tissue homeostasis. Mol Cell Biol. 2009;29(11):3173–3185; doi:10.1128/MCB.01807-08

[SR2] This is adapted and reproduced Figure S5 of the Supplementary material related to Ajima R, Bisson JA, Helt JC, et al. DAAM1 and DAAM2 are co-required for myocardial maturation and sarcomere assembly. Dev Biol. 2015;408(1):126–139. doi:10.1016/j.ydbio.2015.10.003

APPENDIX-C

DETAILS OF STATISTICAL ANALYSIS, PROTEIN-PROTEIN INTERACTION ANALYSIS AND COMMENTARIRES TO CDC5L AND DAAM2