

Выявление общих и уникальных генетических вариантов, ассоциированных с сахарным диабетом 2 типа и гестационным диабетом

Карпова Н.С.*, Дмитренко О.П., Нурбеков М.К.

Научно-исследовательский институт общей патологии и патофизиологии, Москва, Россия

* nataliikarpova.sp@gmail.com

Ключевые слова: полиморфизмы; GWAS; диабет; гестационный сахарный диабет

Мотивация и цель: Гестационный сахарный диабет (ГСД) является наиболее часто встречающейся экстрагенитальной патологией беременности и представляет серьезную медицинскую и социальную проблему, увеличивая частоту нежелательных исходов беременности как для матери, так и для плода [1]. В этиопатогенезе гестационного сахарного диабета выявлен генетический фактор. Poulsen с коллегами (2005) выяснили, что генетические факторы объясняют более 75 % нарушения функции β -клеток поджелудочной железы и гомеостаза глюкозы и 53 % развития инсулинорезистентности [2]. Различные стратегии генетического анализа, такие как общегеномный поиск ассоциации (GWAS), применяются для выявления генетических вариантов (включая однонуклеотидные полиморфизмы (SNP)), связанных с ГСД [3–5]. На сегодняшний день идентифицировано около 40 генетических локусов, ассоциированных с ГСД. Часть из них связана с синтезом и секрецией инсулина (INS, KCNJ11, ABCC8, TCF7L2, ND1), с передачей сигнала инсулина (INSR, IGF2, IRS1) и регулированием углеводного и липидного обмена (PPARG, PPARGC1A, ADRB3, GLUT1, ADIPOQ, FOXC2) [3]. Предыдущие исследования показали, что некоторые генетические локусы, которые предрасполагают к развитию сахарного диабета 2 типа (СД2), также могут предрасполагать к ГСД [6]. Поэтому целью данного исследования был поиск общих и уникальных генетических вариантов, связанных с СД2 и ГСД с последующим анализом отобранных полиморфизмов для обнаружения ключевых генов и путей, вовлеченных в развитие ГСД и СД2.

Методы и алгоритмы: Информацию о полиморфизмах, ассоциированных с ГСД и СД2 получали из базы данных GWAS-catalog (<https://www.ebi.ac.uk/gwas/home>) [3]. Неравновесность сцепления между полиморфизмами оценивали с помощью LDlink (Version 5.6.5, GRCh38). Расположение полиморфизмов относительно области генов осуществляли в Ensembl Variant Effect Predictor (VEP) [7]. Анализ белок-белковых взаимодействий (PPI) проводили с помощью STRING DB (Версия 12.0, <https://string-db.org>) [8]. Из анализа были исключены гены, не кодирующие белки.

Результаты: С 2019 г. были проведены 10 GWAS, в рамках которых обнаружено 30 ассоциаций с ГСД. Из данных полиморфизмов 28 выявлены в рамках одного исследования, в то время как для полиморфизма rs10830963 в гене MTNR1B обнаружена ассоциация с ГСД в нескольких независимых исследованиях (PMID: 36553520, 38182742, 35220425). Согласно LDlink (Version 5.6.5, GRCh38) rs10830962 С аллель коррелирует с аллелем С rs10830963 при анализе всех

популяций ($p < 0.0001$; $D' = 0.9532$; $R^2 = 0.2718$), что свидетельствует в пользу более перспективной для дальнейших исследований ассоциации данного геномного локуса с ГСД.

Мы проанализировали расположение 28 полиморфизмов по отношению к генетическим регионам с помощью Ensembl Variant Effect Predictor (VEP). Большинство полиморфизмов располагается в интронной области гена (67.9 %). Остальные полиморфизмы находятся в межгенной области (10.7 %), области гена (миссенс варианты; 7.1 %), регуляторной области (7.1 %), некодирующего варианта экзона транскрипта (3.6 %), область не определена (3.6 %). Расположение полиморфизмов в регуляторной зоне может означать их влияние на экспрессию генов, образование сплайсинг вариантов генов, а также функционирование регуляторных РНК. Выявлено, что 13 полиморфизмов были ассоциированы как с ГСД, так и с СД2: rs34872471, rs7903146 интрона гена TCF7L2, rs780094 (интрон), rs1260326 (миссенс полиморфизм), rs780093 (интрон) гена GCKR, rs1470579, rs7633675 гена IGF2BP2, rs9348441, rs7754840 интрона гена CDKAL1, rs76895963 в некодирующем экзоне транскрипта CCND2-AS1 и CCND2, rs10811662 в межгенной области CDKN2B-AS1-DMRTA1, rs10830963 интрона гена MTNR1B, rs77464186 интрона гена ARAP1.

Ряд полиморфизмов ассоциирован только с ГСД: rs6798189 гена ADCY5, rs10220124 в межгенной области SLITRK6 – MOB1AP1, rs34499031 интрона гена CDKAL1, rs1402837 интрона гена SPC25, rs1820176 интрона гена MIR583HG и PCSK1, rs6821589 интрон гена PPM1K, rs36090025 интрон гена TCF7L2, rs7254268 в регуляторной области LTBP4-NUMBL, rs10830962 в межгенной области SNRPGP16-MTNR1B, rs9663238 интрона гена HKDC1, rs2274034 интрон гена IL15RA, rs56381411 миссенс вариант гена MAP3K15, rs9275373 в регуляторной области HLA-DQB1-MTCO3P1, rs148031082 в интроне гена PPP1R12A. Следует отметить, что rs36090025 гена TCF7L2 является уникальными для ГСД, а rs34872471, rs7903146 гена TCF7L2 ассоциированы как с СД2, так и с ГСД.

Анализ белок-белковых взаимодействий показал, что для белков, вовлеченных в патогенез ГСД, обогащение было статистически значимо ($p = 4.87 \times 10^{-6}$), однако доказанных физических взаимодействий не было, и только упоминаются вместе в научных публикациях или коэкспрессируются (рис. 1, а). Для белков общих для ГСД и СД2 обогащение было также статистически значимо ($p < 1.0 \times 10^{-16}$), при этом белки DMRTA1, CCND2 и CDKN2B имеют доказанное физическое взаимодействие, а остальные белки упоминаются вместе в научных публикациях или коэкспрессируются (см. рис. 1, б). Все белки, кроме DMRTA1, CCND2 и CDKN2B, вовлечены в развитие ГСД, нарушение углеводного обмена, наследственное нарушение обмена веществ, согласно базе данных DISEASES. Это свидетельствует о том, что данные белки находятся в одном молекулярном маршруте и требуется дальнейшее изучение их роли в развитии ГСД и СД2. Также они могут стать перспективными терапевтическими мишенями для разработки новых лекарственных препаратов.

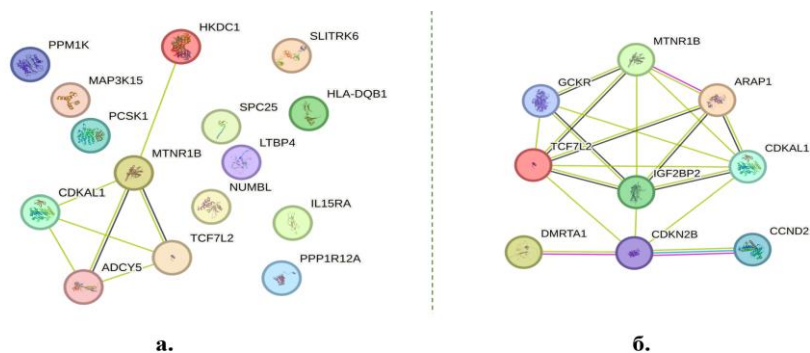


Рис. 1. Визуализация PPI для всех взаимодействия с обозначением белков их взаимодействия. Обозначения доступны в базе данных STRING; а – PPI проведенное для белков в области генов которых располагаются полиморфизмы, ассоциированные с СД2 и ГСД; б – PPI проведенное для белков в области генов которых располагаются полиморфизмы, ассоциированные с СД2 и ГСД

Выводы: Нами были выявлены общие для СД2 и ГСД и уникальные для ГСД ассоциации полиморфизмов. Большинство из 28 проанализированных полиморфизмов являются недавно выявленными и практически неизученными. Расположение полиморфизмов в генах, белковые продукты которых совместно функционируют в общих молекулярных маршрутах создает перспективы как для дальнейшего изучения в рамках исследований случай-контроль, так и для разработки новой таргетной терапии.

Финансирование: Исследование выполнено в рамках государственного задания ФГБНУ «Научно-исследовательский институт патологии и патофизиологии» (№ ФГБНУ-2022-0011: «Выявление значимых биоиндикаторов различных нарушений функций организма»).

Identification of shared and distinct genetic variants associated with type 2 diabetes mellitus and gestational diabetes mellitus

Karpova N.S.*, Dmitrenko O.P., Nurbekov M.K.

Research Institute of General Pathology and Pathophysiology, Moscow, Russia

* nataliakarpova.sp@gmail.com

Key words: polymorphisms; GWAS; diabetes; gestational diabetes mellitus

Motivation and Aim: Gestational diabetes mellitus (GDM) is the most common extragenital pathology of pregnancy and represents a serious medical and social problem, increasing the incidence of adverse pregnancy outcomes for both the mother and the fetus [1]. A genetic factor has been identified in the etiopathogenesis of gestational diabetes mellitus. Poulsen et al. (2005) found that genetic factors explain more than 75 % of the dysfunction of pancreatic β -cells and glucose homeostasis and 53 % of the development of insulin resistance [2]. Various genetic analysis strategies,

such as genome-wide association searches (GWAS), are used to identify genetic variants (including single nucleotide polymorphisms (SNPs)) associated with GDM [3–5]. To date, about 40 genetic loci associated with GDM have been identified. Some of them are associated with the synthesis and secretion of insulin (INS, KCNJ11, ABCC8, TCF7L2, ND1), with insulin signal transmission (INSR, IGF2, IRS1) and regulation of carbohydrate and lipid metabolism (PPARG, PPARGC1A, ADRB3, GLUT1, ADIPOQ, FOXC2) [3]. Previous studies have shown that some genetic loci that predispose to the development of type 2 diabetes mellitus (T2DM) may also predispose to GDM [6]. Therefore, the aim of this study was to search for common and unique genetic variants associated with T2DM and GDM, followed by analysis of selected polymorphisms to discover key genes and pathways involved in the development of GDM and T2DM.

Methods and Algorithms: Information on polymorphisms associated with GDM and T2DM was obtained from the GWAS-catalog database (<https://www.ebi.ac.uk/gwas/home>) [3]. Linkage disequilibrium between polymorphisms was assessed using LDlink (Version 5.6.5, GRCh38). The location of polymorphisms relative to the gene region was carried out in Ensembl Variant Effect Predictor (VEP) [7]. Protein-protein interaction (PPI) analysis was performed using STRING DB (Version 12.0, <https://string-db.org>) [8]. Genes that do not code for proteins were excluded from the analysis.

Results: Since 2019, 10 GWAS have been conducted and found 30 associations with GDM. Of these polymorphisms, 28 were identified in a single study, while the rs10830963 polymorphism in the MTNR1B gene was found to be associated with GDM in several independent studies (PMID: 36553520, 38182742, 35220425). According to LDlink (Version 5.6.5, GRCh38), the rs10830962(C) allele correlates with the rs10830963(C) allele when analyzing all populations ($p < 0.0001$; $D' = 0.9532$; $R^2 = 0.2718$), which indicates the benefit of more promising for further studies of the association of this genomic locus with GDM.

We analyzed the location of 28 polymorphisms relative to genetic regions using the Ensembl Variant Effect Predictor (VEP). The majority of polymorphisms are located in the intronic region of the gene (67.9 %). The remaining polymorphisms are located in the intergenic region (10.7 %), the gene region (missense variants; 7.1 %), the regulatory region (7.1 %), the non-coding variant of the transcript exon (3.6 %), the region is not defined (3.6 %). The location of polymorphisms in the regulatory zone may indicate their influence on gene expression, the formation of splicing variants of genes, as well as the functioning of regulatory RNAs. It was revealed that 13 polymorphisms were associated with both GDM and T2DM: It was revealed that 13 polymorphisms were associated with both GDM and T2DM: rs34872471, rs7903146 of the TCF7L2 gene intron, rs780094 (intron), rs1260326 (missense polymorphism), rs780093 (intron) GCKR gene, rs1470579, rs7633675 of the IGF2BP2 gene, rs9348441, rs7754840 intron of the CDKAL1 gene, rs76895963 in the non-coding exon of the CCND2-AS1 and CCND2 transcript, rs10811662 in the intergenic region of CDKN2B-AS1-DMRTA1, rs10830963 intron of the MTNR1B gene, rs77464186 of the intron of the ARAP1 gene. A number of polymorphisms are associated only with GDM: rs6798189 of the ADCY5 gene, rs10220124 in the intergenic region of SLITRK6 – MOB1AP1, rs34499031 of the intron of the CDKAL1 gene, rs1402837 of the intron of the SPC25 gene, rs1820176 of the intron of the MIR583HG and PCSK1 gene, 589 intron of the PPM1K gene, rs36090025 intron of the TCF7L2 gene, rs7254268 in the regulatory region LTBP4-

NUMBL, rs10830962 in the intergenic region of SNRPGP16-MTNR1B, rs9663238 intron of the HKDC1 gene, rs2274034 intron of the IL15RA gene, rs56381411 missense variant of the MAP3K15 gene, rs9275373 in the regulatory region of HLA-DQB1-MTCO3P1, 8031082 in the intron of the PPP1R12A gene. It should be noted that rs36090025 of the TCF7L2 gene is unique to GDM, and rs34872471, rs7903146 of the TCF7L2 gene are associated with both T2DM and GDM.

Analysis of PPI showed that for proteins involved in the pathogenesis of GDM, the enrichment was statistically significant ($p = 4.87\text{e-}06$), but there were no proven physical interactions. and are only mentioned together in scientific publications or co-expressed (Fig. 1a). For proteins common to GDM and T2DM, the enrichment was also statistically significant ($p < 1.0\text{e-}16$), with the proteins DMRTA1, CCND2 and CDKN2B having a proven physical interaction, and the remaining proteins being mentioned together in scientific publications or being co-expressed (Fig. 1b). All proteins except DMRTA1, CCND2 and CDKN2B are involved in the development of GDM, carbohydrate metabolism disorders, and inherited metabolic disorders, according to the DISEASES database. This indicates that these proteins are located in the same molecular route and further study of their role in the development of GDM and T2DM is required. They can also become promising therapeutic targets for the development of new drugs.

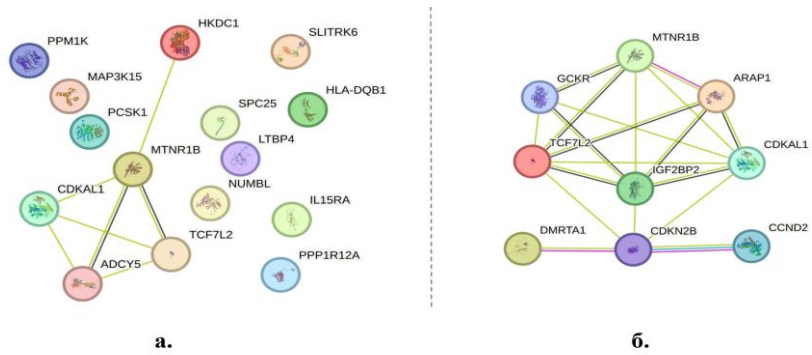


Fig. 1. Visualization of PPIs for all interactions with their interaction proteins labeled. Designations are available in the STRING database; a. PPI carried out for proteins in the gene region of which polymorphisms associated with T2DM and GDM are located; b. PPI carried out for proteins in the gene region of which polymorphisms associated with T2DM and GDM are located

Conclusion: We identified associations of polymorphisms common to T2DM and GDM and unique to GDM. Most of the 28 analyzed polymorphisms are recently identified and practically unstudied. The location of polymorphisms in genes whose protein products function together in common molecular pathways offers promise for both further study in case-control studies and for the development of new targeted therapies.

Funding: The study is supported within the framework of the state assignment of the Federal State Budgetary Institution “Research Institute of Pathology and Pathophysiology” (No. FGFU-2022-0011: “Identification of significant bioindicators of various disorders of body functions”).

Список литературы/References

1. Damm P. et al. Gestational diabetes mellitus and long-term consequences for mother and offspring: a view from Denmark. *Diabetologia*. 2016;59:1396-1399
2. Poulsen P. et al. Heritability of insulin secretion, peripheral and hepatic insulin action, and intracellular glucose partitioning in young and old Danish twins. *Diabetes*. 2005;54(1):275-283
3. Sollis E. et al. The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. *Nucleic Acids Res*. 2023;51(D1):D977-D985
4. Dias S. et al. Molecular biomarkers for gestational diabetes mellitus. *Int J Mol Sci*. 2018;19(10):2926
5. Lu W. et al. Molecular biomarkers for gestational diabetes mellitus and postpartum diabetes. *Chin Med J*. 2022;135(16):1940-1951
6. Kawai V.K. et al. A genetic risk score that includes common type 2 diabetes risk variants is associated with gestational diabetes. *Clin Endocrinol*. 2017;87(2):149-155
7. McLaren W. et al. The ensembl variant effect predictor. *Genome Biol* 2016;17(1):122
8. Szklarczyk D. et al. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res*. 2023;51(D1):D638-D646. doi 10.1093/nar/gkac1000