

UDC 577.218

Creation of gene expression database on preeclampsia-affected human placenta

O. Lykhenko, A. O. Frolova, M. Yu. Obolenskaya

Institute of Molecular Biology and Genetics, NAS of Ukraine
150, Akademika Zabolotnoho Str., Kyiv, Ukraine, 03680
lykhenko.olexandr@gmail.com

Aim. Publication of gene expression raw data in public repositories made it possible to reuse these data for cross-experiment integrative analysis and make new insights into biological phenomena. However, data uploaded by independent contributors are not standardized, sometimes incomplete and need preprocessing before any further analysis. Hence, the aim of this study was to create a specialized database of gene expression profiles, particularly in preeclampsia-affected human placenta as a cause of high rate of morbidity and mortality all over the world with unknown etiology and pathogenesis. **Methods.** All experiment and sample metadata were automatically extracted from ArrayExpress database via Bioservices. NCBI database was used to supplement the missing data along with the corresponding scientific articles and authors personal data. The experimental sample attributes were standardized according to MeSH term dictionary and Experimental Factor Ontology. **Results.** A database of more than 1000 samples of normal and preeclampsia-affected human placenta was created and supplied with metadata containing information on biological specimen, diagnosis, gestational age, mode of delivery and other sample characteristics. **Conclusion.** The samples in our newly created database now contain metadata for them to be comparable. The biological samples may be arranged in different case-control groups of larger size than in individual datasets for statistically significant analysis.

Key words: Genetic Databases; Placenta; Preeclampsia; Oligonucleotide Array; Sequence Analysis; Microarray Data

Background

Open gene expression databases have over time acquired tremendous amounts of data [1, 2]. It is now possible to make original discoveries just by analyzing these data. An *integrative analysis* is one of the approaches in this

case. Unlike *meta-analysis*, which is essentially attaching the results of different experiments, an integrative analysis implies merging or integration of raw data which, as studies show [3, 4], identifies significantly more differentially expressed genes. Besides, integrative analysis provides larger sample sets and

© 2017 O. Lykhenko *et al.*; Published by the Institute of Molecular Biology and Genetics, NAS of Ukraine on behalf of Biopolymers and Cell. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited

thus increases statistical significance and reduces experimental bias [5], which makes it most useful in cases when individual average sample set is small.

To identify which pieces of data are similar enough for the integration to make sense each individual sample must be supplemented with at least minimal data. Here and later these sample clinical and biological information will be called sample *metadata* (do not confuse with meta-analysis, a method to unite results of different studies). Unfortunately, most popular public databases like NCBI or ArrayExpress do not always contain sufficient sample-wise metadata. There are several reasons for that. The main one is that the integration is not a primary goal for these databases. As is stated in [6] the ArrayExpress primary goal is to serve as an archive for microarray data associated with scientific publications and other research. Also, since the data were uploaded by independent contributors the biological sample metadata is not standardized and, consequently, not ready for immediate automated access by a search query. The choice of relevant characteristics of biological sample is also up to a contributor and those characteristics do not always match the needed ones for the data integration.

To address these problems investigators develop specialized databases. Here we present a curated collection of publicly available datasets on preeclampsia-affected human placenta.

Preeclampsia is a disorder that occurs only during pregnancy and the postpartum period and affects both the mother and the unborn baby. Affecting at least 5–8 % of all pregnancies, it is a rapidly progressive condition cha-

racterized by high blood pressure and the presence of protein in the urine. Typically, preeclampsia occurs after 20 weeks gestation (in the late 2nd or 3rd trimester or middle to late pregnancy) and up to six weeks postpartum (after delivery), though in rare cases it can occur earlier than 20 weeks. Globally, preeclampsia and other hypertensive disorders of pregnancy are a leading cause of maternal and infant illness and death.

Although etiology and pathogenesis of preeclampsia are still unknown, numerous studies (listed in Supplement 1) point at the gene dysregulation in placenta to be one of the possible prerequisites for the disease. Besides, there are also known cases when placenta develops without fetus, called molar, which are associated with very early-onset preeclampsia. This is why we focused our study on the gene expression in placenta. Finally, we chose cDNA microarray technology as most popular in preeclampsia studies among ones providing information on the whole transcriptome.

Our database now contains sufficient metadata for the samples to be united into relevant case-control groups for the subsequent integrative analysis and further search for differentially expressed genes and inferring gene networks.

Methods

Our software is written in Python language using Django framework for web interface development and Postgres for relational database support. All experiment and sample metadata were automatically extracted from ArrayExpress database via Bioservices which is a Python interface to ArrayExpress. NCBI

database was used to supplement the missing data along with the corresponding scientific articles and authors personally.

Source code and database backup file are available at GitHub:

<https://github.com/Sashkow/placenta-pre-eclampsia>

Web interface for our database at its current stage can be accessed at:

<http://194.44.31.241:24173/>

Results

The database at its current stage contains a total of 32 experiment datasets and more than 1000 placenta samples, about 900 of which contain minimal metadata for relevant cross-experiment study groups to be constructed, which is diagnosis, gestational age and tissue type. A separate group of 11 experiment datasets and 300 samples are in vitro experiments with cell cultures and immortalized cell lines as biological samples. Apart from preeclampsia, some related complication were considered including fetal growth retardation, HELLP syndrome and some genetic disorders such as mosaicism and trisomy of chromosome 16.

Database structure

Figure 1 shows general structure of the database. Rectangles represent tables in database. Tables are linked with one-to-many and many-to-many relations. For example, Experiments table is in many-to-many relation with Microarrays table for each experiment may utilize multiple microarray platforms and each platform can be used in multiple experiments. On the other hand, Experiments table is in one-to-many relation with Samples table as experiment may contain multiple samples while each sample can be a part of only *one* experiment. Experiments and Microarrays tables contain HStore field for storing mappings of strings to strings which are experiment/microarray attribute name-value pairs like “accession:E-GEOD-42424” in our case. SampleAttributes table containing all attributes for all samples is here to provide opportunity for different samples to have different attributes and to make new names reverse compatible with the originally downloaded ones. StandardSampleAttributeName and StabdarSampleAttributeValues tables store standard terms for sample attribute names and values. The field named

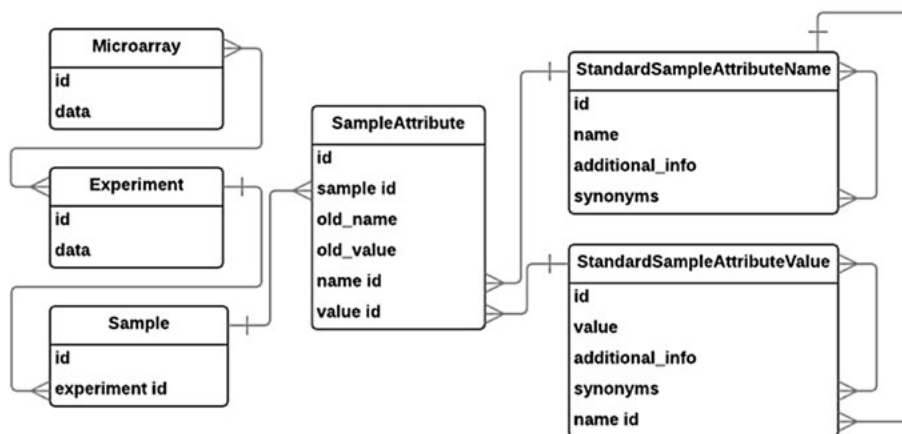


Fig. 1. Database entity relation diagram

additional_info is HStore field with information about such as term source and short description. These tables for standard terms also contain *synonyms* many-to-many fields with themselves providing lists of synonymous terms. Detailed explanation of the database design process can be found in Supplement 3.

Choice of datasets

The initial list of 43 relevant datasets was obtained as a result of ArrayExpress search by the following query: “preeclampsia OR preeclampsia OR preeclamptic OR preeclamptic” with results filtered by organism “Homo sapiens”, experiment type “rna assay”, experiment type “array assay”. E-GEOD-25906 was excluded due to data retrieval failure. E-MTAB-3732 was excluded since it is a compilation of microarray experiments for different diseases taken from publicly available sources. E-GEOD-15787, E-GEOD-22526, E-MEXP-1050 were excluded due to old microarray design or failure to find probe nucleotide sequences for the array. The full list of included and excluded ones can be found in Supplement 1.

Sample attribute names and values standardization

Medical Subject Headings (MeSH) terms dictionary was chosen as a standard for naming biological sample attribute names and values. We also used Ontology Lookup Service (OLS) as a secondary source in cases where no fitting MeSH term was found, since OLS utilizes multiple ontologies and has been recently updated to have more convenient interface than it used to. Furthermore, ArrayExpress itself uses one of OLS’s ontologies, namely

Experimental Factor Ontology (EFO), to perform advanced search over genetic experiments and biological samples. It means that our standardized samples metadata is potentially compatible with ArrayExpress and might be used to improve its search quality regarding datasets we have processed.

Here comes the list of standard sample attribute names and values, which are MeSH or other ontology terms, with lists of mappings onto sample attribute names and values originally downloaded from ArrayExpress. The format is the following:

```
Standard Name 1 (list of original names)
  Standard Value 1 (list of original values)
  Standard Value 2 (list of original values)
  ...
Standard Name 2 (list of original names)
  Standard Value 1 (list of original values)
  Standard Value 2 (list of original values)
  ...
```

Note that mapping is ambiguous meaning that an original name may map onto different standard names depending on the context of the experiment. For example, original name “placenta” will map onto standard name “Chorion” if the context of the experiment suggests that. Some standard names do not map on any original name since that information was absent in ArrayExpress data and obtained manually either from the corresponding articles or from the authors personally. Here is a list of mappings for Diagnosis sample attribute provided as an example. The entire list of mappings can be found in Supplement 2.

```
Diagnosis (diagnosis, subject status, disease, disease state, disease status, genotype, condition,
```

group, phenotype, classification, DiseaseState, Disease State)
Healthy (normotensive, normotensive control patient, normal, <empty>, healthy, control, Health, preterm labor, Normotensive control pregnancy, preterm)
Pre-Eclampsia (preeclamptic, pre-eclamptic patient, pre-eclampsia, Preeclampsia, EOPET, PE, Preeclampsia after 20 weeks of gestation to 140/90 mm Hg, excretion of 0.3 g protein in 2 urine samples, early onset preeclampsia, late onset preeclampsia, mild preeclampsia)
Pre-Eclampsia and HELLP
Pre-Eclampsia and FGR (EOPET, pre-eclampsia and fgr, preeclampsia+infant SGA)
Severe Preeclampsia (severe preeclampsia, preeclampsia, severe preeclampsia)
Fetal Growth Retardation (IUGR, Fetal growth restriction (FGR), infant SGA)

Sample metadata completeness

While sample metadata from all considered datasets are now *standardized*, sample metadata remains far from complete. Yet it is already sufficient to tell whether any two of the samples are *comparable* i.e. whether they can be put to the same case-control group or not. We identified three key attributes that shall be our criteria for comparability: Diagnosis, Gestational Age, Biological Specimen – only the samples of the same tissue at the same

stage of the placental development under similar conditions can be put into the same study group whenever search for differentially expressed genes between control and deviation groups is to be performed.

Gestational Age for some reason is often not mentioned directly in sample data. However, it is possible to infer a meaningful category for gestational age by looking, for instance, at whether delivery was performed at term (37 to 41 week of gestation) or prematurely (20 to 37 weeks) or at pregnancy trimester if mentioned. Also, authors often publish gestational age mean and deviation for a study group which as well can be used to determine gestational age category for the samples. We used Gestational Age distribution for samples with known exact value to identify five gestation categories: first trimester, second trimester, early preterm, late preterm, term (Figure 3). Then we used approximate gestational age information for the samples with unknown exact Gestational Age value to put them into one of the five suggested gestation categories. Thus we determined gestational age for all samples within gestational age category.

Noteworthy, the value of gestational age depends on a measurement technique. If it is measured as first day of the last menstrual cycle it is roughly 2 weeks earlier than if measured since actual conception date.

Although **Diagnosis** is specified almost everywhere as the majority of the considered experiments are case-control in design, it is not always enough just to know value of this attribute. For example, if the sample is a “control” it can be placenta delivered from natural birth or after c-section, it can be term or preterm birth, medically induced or spontaneous. All

these characteristics are to be considered when matching controls to cases during formation of cross-experiment study groups. Likewise, when the sample is “preeclampsia” there are also several additional attributes to look at: preeclampsia onset, severity, additional complications, formal preeclampsia criteria such as blood pressure and urine protein concentration.

Biological Specimen is mostly placenta in our study. A biological sample tissue type is specified wherever mentioned to reduce the noise that would be caused by differential expression among the cell types inside placenta.

The sample attribute coverage summary for the rest of attributes are in Figure 2. Metadata may be further completed in future either through utilizing existing approximate data as was done here with gestational age or by inferring metadata directly from the expression data as was done, for instance, with fetal sex [7].

Most of the considered samples are comparable and ready for integrative analysis at this moment. These biological samples can now

constitute the case-control groups of larger size than in the original datasets.

Discussion

Numerous attempts to organize the gene expression data for cross-experiment analysis have been undertaken by different investigators. The integrative approach for preeclampsia study is published [7], though no database creation was intended. The authors merged 7 gene expression microarray datasets of case-control studies of preeclampsia affected placenta and created study groups of 77 and 96 samples for preeclampsia and control groups, respectively. The authors got interesting insights into the nature of the disease that could not have been made from mere observation of individual studies results. Namely, unsupervised clustering based on both sample expression and metadata revealed three clusters of samples. One of the clusters turned out to be largely composed of the healthiest of the healthy term placentas along with supposedly

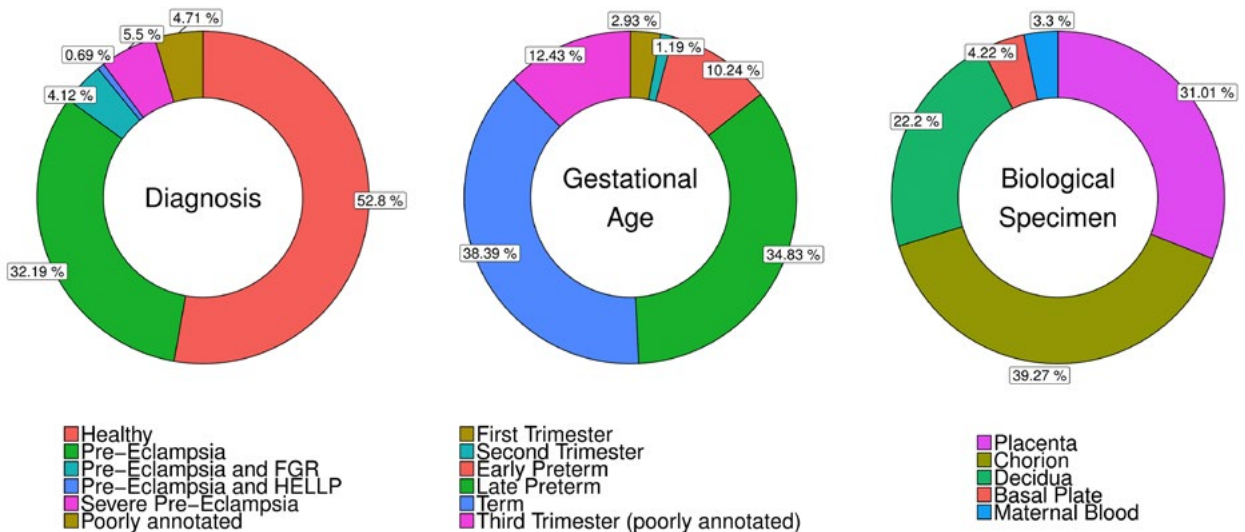


Fig. 2. Sample core characteristics.

preeclamptic ones [7]. A hypothesis was proposed that unhealthy placentas of that cluster had been misdiagnosed as preeclamptic ones, and were really afflicted with another maternal hypertensive disorder, such as gestational hypertension or chronic hypertension. This is supported by the GSEA and gestational age comparisons, which indicate that cluster 1 is largely composed of the healthiest term placentas in this data set.

GENEVESTIGATOR is a search engine for gene expression over a compendium of manually curated datasets [8, 9]. It provides variety of tools addressing the questions of finding conditions under which the genes of interest are the most up- or down-regulated as well as finding the genes that are the most expressed under certain conditions. Biological samples are carefully annotated using custom ontology with a variety of characteristics allowing to filter for highly specific cases. Cross-experiment analysis is based on a concept of meta-profile. As explained in [10] meta-profiles summarize the expression levels from many samples according to their biological context. Each sample is annotated with five attributes: anatomical parts, cell lines, cancer types (for cancer samples), developmental stages, and perturbations – to generate meta-profiles. Hence each meta-profile is an expression profile of an “average” sample generated from all expression profiles of the samples of specific tissue or developmental stage. An exception is the Perturbation meta-profile, which consists of responses to various experimental conditions (drugs, chemicals, hormones, *etc.*), diseases, and genotypes. For Perturbation the meta-profile results are created by comparing groups of samples from

individual experiments. Data from multiple experiments are not mixed to create a single value. As a result, this tool contains large compendia of response types collected from many experiments. It also worth mentioning that samples of two different microarray platforms cannot be considered simultaneously.

This implies that **GENEVESTIGATOR** is closer to meta-analysis than to integrative analysis. It integrates the sample data for samples of the same platform and biological context to create meta-profiles but other tasks such as comparison of the samples of different biological contexts or comparison of differentially expressed genes for different medical or biological cases (perturbations) are pure meta-analysis.

The **GENEVESTIGATOR** does not currently contain datasets of our interest and does not allow user to upload data directly and, thus, cannot satisfy our goals.

Another resource enabling interactive query and navigation of transcriptome datasets is **Gene Expression Browser** [11] and its specific implementation for the placental gene expression [12]. While its data is more relevant to our pursues and the interface is quite convenient and has some extended features in comparison to GEO or ArrayExpress, such as search for differentially expressed genes in a single experiment, building relevant plots, giving detailed information on found genes and a bit of meta-analysis tools such as search for the experiments that have given gene differentially expressed with a certain rate of difference, no tools for integrative analysis are provided.

An Integrative Meta-Analysis of Expression Data (**INMEX**) service [13] is the one having options for both meta- and integrative analysis, not to mention a huge ensemble of tools for

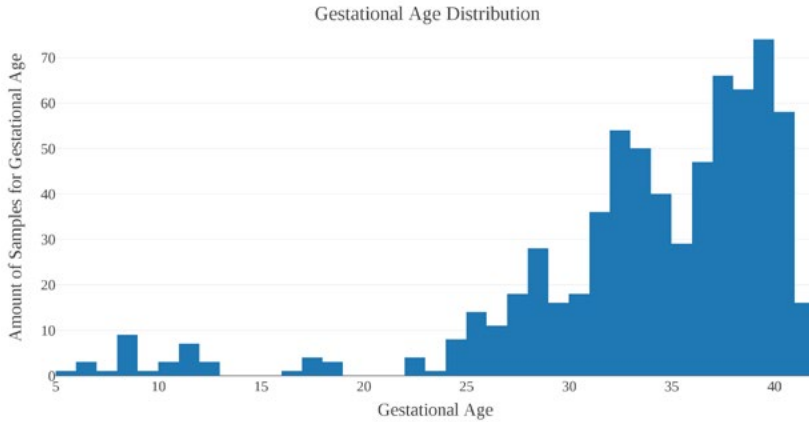


Fig. 3. Gestational age distribution for samples with known exact value. Five gestation categories may be considered: first trimester, second trimester early preterm (25 to 30 week), late preterm (30 to 37 week), term (37 to 42 week).

further data analysis. It is worth emphasizing that INMEX is not a database, but a pure tool with no pre-uploaded data except for several examples: uploading raw data and constructing proper study groups are the user responsibility. Opportunely, our newly developed database can make for a perfect data input to INMEX and it is quite possible that INMEX's output will be a solution to pursue our data integration and subsequent analysis.

Finally, we cannot but emphasize how important it is to upload well annotated *sample-wise* experiment metadata. Doing so we join the call from Drs. D.M. Nelson and G.J. Burton, editors of the Placenta journal, who published

a technical note to improve the reporting of studies of the human placenta featuring those essential parameters attention to which “will lead to enhanced chances that comparisons will be of apples with apples, instead of comparisons of two fruits” [14]. While we notice the improvement in time of the data reporting quality we could not manage to gather even the half of what the note suggests from the publicly available datasets and articles (Figure 4). Namely, the information regarding drugs, previous prenatal admissions, screened for diabetes, antibiotics, beta strep status, antenatal steroids, magnesium sulfate, anesthesia and cervical ripening agent was almost entirely absent.

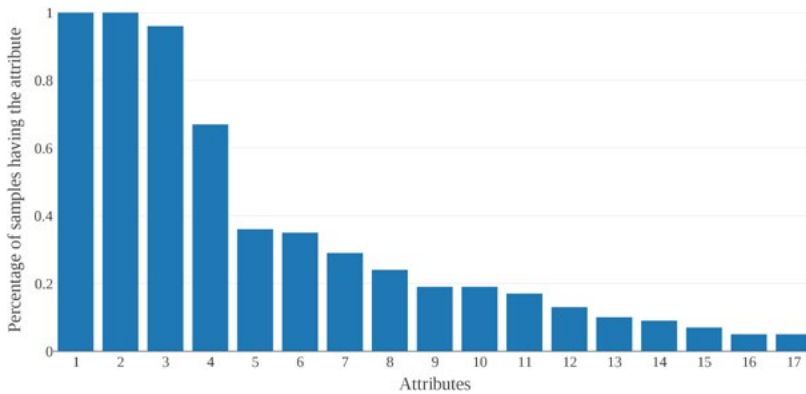


Fig. 4. Percentage of samples containing attribute.

Conclusions

Most of the samples in our database are now provided with sufficient metadata for them to be comparable. These biological samples can now constitute case-control groups of larger size than in individual datasets. Described here data gathering and standardization are the first steps to further analysis, namely, the search for differentially expressed genes between different cross-experiment case and control groups. Also, our database is designed and expected to be enlarged with forthcoming gene expression experiments metadata and, potentially, with data from other omics (genome, metabolome) with a relatively low effort.

Competing interests

The authors declare that they have no competing interests.

Additional Files

Supplement 1 — ArrayExpress Experiments' Accession Numbers

List of ArrayExpress accession numbers for experiments taken for the study.

Supplement 2 — Sample Attribute Names and values

List of all standardized names and values for sample attributes along with the corresponding names and values originally downloaded from ArrayExpress.

REFERENCES

1. Barrett T, Troup DB, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Muetter RN, Holko M, Ayanbule O, Yefanov A, Soboleva A. NCBI GEO: archive for functional genomics data sets — 10 years on. *Nucleic Acids Res.* 2011;**39**(Database issue): D1005-10.
2. Rustici G, Kolesnikov N, Brandizi M, Burdett T, Dylag M, Emam I, Farne A, Hastings E, Ison J, Keays M, Kurbatova N, Malone J, Mani R, Mupo A, Pedro Pereira R, Pilicheva E, Rung J, Sharma A, Tang YA, Ternent T, Tikhonov A, Welter D, Williams E, Brazma A, Parkinson H, Sarkans U. ArrayExpress update—trends in database growth and links to data analysis tools. *Nucleic Acids Res.* 2013;**41**(Database issue):D987-90.
3. Frolova A, Obolenska M. Integrative Approaches for Data Analysis in Systems Biology: Current Advances. In: Proceedings of Applied Physics and Engineering (YSF), 2016 II International Young Scientists Forum:194–98.
4. Taminau J, Lazar C, Meganck S, Nowé A. Comparison of Merging and Meta-Analysis as Alternative Approaches for Integrative Gene Expression Analysis. *ISRN Bioinform.* 2014(2014): 345106.
5. Walsh CJ, Hu P, Batt J, Santos CC. Microarray Meta-Analysis and Cross-Platform Normalization: Integrative Genomics for Robust Biomarker Discovery. *Microarrays (Basel).* 2015;**4**(3):389–406.
6. Sarkans U, Parkinson H, Lara GG, Oezcimen A, Sharma A, Abeygunawardena N, Contrino S, Holloway E, Rocca-Serra P, Mukherjee G, Shojatalab M, Kapushesky M, Sansone S-A, Farne A, Rayner T, Brazma A. The ArrayExpress gene expression database: A software engineering and implementation perspective. *Bioinformatics.* 2005;**21**(8):1495–501.
7. Leavey K, Bainbridge SA, Cox BJ. Large scale aggregate microarray analysis reveals three distinct molecular subclasses of human preeclampsia. *PLoS One.* 2015;**10**(2):e0116508.
8. Hruz T, Laule O, Szabo G, Wessendorp F, Bleuler S, Oertle L, Widmayer P, Gruissem W, Zimmermann P. GENVESTIGATOR V.3: a reference expression database for the meta-analysis of transcriptomes. *Adv Bioinformatics.* 2008; (2008):420747.
9. Zimmermann P, Hirsch-Hoffmann M, Hennig L, Gruissem W. GENESTIGATOR. Arabidopsis

- microarray database and analysis toolbox. *Plant Physiol.* 2004;**136**(1):2621–32.
10. Clauzel C, Foltête JC, Girardet X and Vuidel G. GENEVESTIGATOR, a High Performance Search Engine for Gene Expression. *User Manual.* 2016; no. May: 0–37.
 11. Speake C, Presnell S, Domico K, Zeitner B, Bjork A, Anderson D, Mason MJ, Whalen E, Vargas O, Popov D, Rinchai D, Jourde-Chiche N, Chiche L, Quinn C, Chaussabel D. An interactive web application for the dissemination of human systems immunology data. *J Transl Med.* 2015;**13**:196.
 12. Marr AK, Boughorbel S, Presnell S, Quinn C, Chaussabel D, Kino T. A curated transcriptome dataset collection to investigate the development and differentiation of the human placenta and its associated pathologies. *Version 2. F1000Res.* 2016 [revised 2016];**5**:305.
 13. Xia J, Fjell CD, Mayer ML, Pena OM, Wishart DS, Hancock RE. INMEX--a web-based tool for integrative meta-analysis of expression data. *Nucleic Acids Res.* 2013;**41**(Web Server issue):W63–70.
 14. Nelson DM, Burton GJ. A technical note to improve the reporting of studies of the human placenta. *Placenta.* 2011;**32**(2):195–6.

Розробка бази даних генної експресії плаценти людини в умовах преєклампсії

А. Лихенко, А. Фролова, М. Ю. Оболенська

Мета. Публікація «сырих» результатів у відкритих репозиторіях надає можливість поєднати дані декількох експериментів для інтегративного аналізу, який сприятиме підвищенню статистичної значущості висновків, а відтак і кращому розумінню біологічних явищ. Однак, оскільки дані у відкритих репозиторіях були завантажені незалежними дослідниками за відсутності єдиних стандартів, то назви характеристик зразків (атрибути) виявилися не уніфікованими, часто неповними та потребують додаткової обробки, перш ніж інтеграція стане можливою. **Метою** роботи є розробка спеціалізованої бази даних генної експресії в плаценті людини в умовах преєклампсії як причини високої захворюваності та смертності по всьому світу за невідомої етіології захворювання та його патогене-

зу. **Методи.** Метадані експериментів та біологічних зразків були автоматично завантажені з ArrayExpress за допомогою Bioservices. Відсутні дані були отримані з бази NCBI, а також із текстів відповідних статей та внаслідок спілкування з авторами статей. Метадані були стандартизовані за допомогою словника MeSH та Experimental Factor Ontology. **Результати.** База даних з більш як 1000 зразків плаценти в нормі та за умов преєклампсії була розроблена і доповнена метаданими, які включають характеристику біологічних зразків, клінічний діагноз, вік жінки, стать плоду, характеристику пологів та інші допоміжні ознаки зразків. **Висновки.** Біологічні зразки розробленої бази даних охарактеризовані максимально можливою кількістю метаданих, що дозволяє створювати різні групи порівняння профілів генної експресії. Об'єднання профілів генної експресії (інтеграція) різних зразків на підставі ідентичності метаданих з наступним аналізом об'єднаних даних підвищить статистичну значущість висновків в порівнянні з аналізом результатів окремо взятих дослідів.

Ключові слова: генетичні бази даних, плацента, преєклампсія, мікромасив олігонуклеотидів, аналіз послідовностей, дані мікромасивів.

Разработка базы данных генной экспрессии плаценты человека в условиях преэклампсии

А. Лихенко, А. Фролова, М. Ю. Оболенская

Цель. Публикация «сырых» результатов в публичных репозиториях открывает возможность объединения данных нескольких экспериментов для интегративного анализа, который способен улучшить понимание биологических явлений. Однако, поскольку данные в публичных репозиториях были загружены независимыми исследователями при отсутствии общих стандартов, многие характеристики образцов оказались не унифицированными, часто неполными и требуют дополнительной обработки, прежде, чем интеграция данных станет возможной. **Целью** данной работы было создание специализированной базы данных генной экспрессии в плаценте человека в условиях преэклампсии как причины высокой заболеваемости и смертности беременных во всем мире при неизвестной этиологии заболевания и его патогенеза. **Методы.**

Метаданные экспериментов и биологических образцов были автоматически загружены из ArrayExpress с помощью Bioservices. Недостающие данные были получены из базы данных NCBI, а также — из текстов соответствующих статей и в результате общения с авторами. Атрибуты образцов экспериментов были стандартизированы с помощью словаря медицинских терминов MeSH и Experimental Factor Ontology. **Результаты.** База данных из более, чем 1000 образцов плаценты человека в норме и при прееклампсии, была разработана и дополнена метадаанными, включающими характеристику биологических образцов, клинический диагноз, возраст женщины, пол плода, особенности родоразрешения и другие вспомогательные характеристики. **Выводы.** Биологические образцы разрабо-

танной базы данных охарактеризованы максимально возможным количеством метадаанных, что позволяет формировать разные группы сравнения профилей генной экспрессии. Объединение профилей генной экспрессии (интеграция) разных образцов на основе идентичных метадаанных с последующим анализом объединенных данных повысит статистическую надежность выводов по сравнению с анализом данных отдельно взятых экспериментов.

Ключевые слова: генетические базы данных; плацента; прееклампсия; микромассив олигонуклеотидов; анализ последовательности; данные микромассивов.

Received 20.05.2017