7. Sekov I.N., Afanaseva D.N. Index of a teta/beta rhythms capacity during a beta-based neurobiofeedback at children with ADHD//Yakut medical journal. - 2008. №3.- 31 p.
8. Chuhrova V. A clinical electroencephalography (the grant for doctors). - M.:1990.–12-15 p.
9. Shtark of M. B. Electroencephalographic neurobiofeedback at a syndrome of deficiency of attention with hyperactivity [add/hd a syndrome - a harbinger addict disorders] / M.B.Shtark, O.A.Dzhafarova, A.B.Skok, E.V.Hajmovich, O.S.Shubina//Narcology, 2004. - 56-62 p.
10. American Psychiatric Association. (2000). Diagnostic and statistical manual of mental disorders (4th ed., text revision).-Washington, DC: Author: Ariel, R., Dunlosky, J., & Bailey, H. (2009). Agenda-based regulation of study-time allocation: When agendas override item-based monitoring. *Journal of Experimental Psychology: General, 138, 432–447.*

Development of data base`s structure for biobanks.

Buikin S.V., Bragina E.Y., Koneva L.A.

Institute of Medical genetics Siberian Branch of Russian Academy of Medical Sciences, Tomsk, 634050 fax:(3822)51-37-44;* e-mail: stepan.buikin@medgenetics.ru

Summary: Collection, preprocessing and use of biological samples cause many organizational problems, which are successfully resolved with help of biobanks. Among the components of successful work of biobank is a practical, functional and reliable database for storage of information. In the paper, principals and problems of database development for biobanks are considered, taking example the biobank of the Institute of Medical genetics SB RAMS.

Keywords: biobank, database, multifactorial diseases, genes.

Introduction: The intensification of genetic studies of the basis of complex diseases has allowed to formulate basic propositions concerning study design, and also concerning qualitative and quantitative composition of the groups under investigation [8, 10, 12, 13]. Requirements to the size, uniformity and description of the samples become more strict [1, 7], and this leads to

substantial growth of time and fund expenses on registration, standardization and preservation of biological samples and the accompanying information. Biobanks can be optimal decision of the given problem because they consolidate large arrays of biological samples and the accompanying information [2]. One of the basic components of successful work of biobanks is the practic, functional and reliable databases (DB) for storage of the accompanying information. In the article, main principles and problems of database development for biocollection resources are considered, taking as an example the DB for the Biobank of the Institute of Medical Genetics of Russian Academy of Medical Science.

DB Designing is one of the most difficult and important tasks in the process of creation of an information system. As a result of the designing, the content of the DB, data structure and management tools should be established. Process of a DB designing includes the following stages: (1) conceptual or information-logical designing; (2) logic designing; (3) physical designing [4]. The information-logical model represents the description of structure and dynamics of a subject domain, information requirements of the system users given in terms which are clear to the user and independent of realization of the system. Thus, at this stage are determined: the basic objects of a subject domain (objects about which the information should be kept in a DB), attributes of objects, connections between objects, typical queries to the DB. Logic designing means the decision on and optimum representation of objects from subject domain in abstract objects of the data model given that this representation does not contradict semantics of the subject domain. Physical designing should ensure efficiency of performance of queries to the DB [5]. All stages should consider specific character of a DB under development.

Biobank of the Institute of Medical Genetics of Siberian Branch of Russian Academy of Medical Sciences is designed to store unique biological material with a detailed description of the genetic, demographic and clinical information. The Biobank contains several thousands of DNA samples and blood samples of individuals with various pathologies, including cardiovascular, immune, infectious, monogenic and other diseases as well as DNA of individuals from different populations of Eastern and Western Siberia and Central Asia [2]. Related information on the samples include: (1) description of medical history; (2) the results of physical and laboratory examination of individuals; (3) data on large number of genetic markers (the results of genotyping, sequencing, karyotyping, etc.); (4) ethical issues on use of the samples for various studies (individual informed consent, date and number of the conclusion of the Ethics Committee); (5) Lab working details (a method of DNA extraction, concentration, quality, sample volume, etc.). Thus, the accompanying information for the biological material includes

diverse data for different pathologies, and is stored in the form of Excel tables, text and graphic files.

Formalization of biomedical information that characterizes the biological samples is a complex task. There are strict requirements to the collecting and storing of the information which allow to minimize errors and inaccuracies because such information is crucial for assessing the association of the particular genotype (polymorphic variant of a gene) with the development of the disease [9].

The aim of developing a database for the Biobank RIMG is to ensure data integrity and efficiency of scientific research. The entire data set was structured according the goals and objectives of the Biobank RIMG. It should be noted that not only the current needs of researchers have been taken into account but also the possibility of using the biological material and information for biomedical studies in future. In addition, we take into account the ethics issues, and protecting of intellectual property associated with the storage and presentation of the results for the biological samples from the biobanks [11].

Database of the Biobank of Research Institute of Medical Genetics (RIMG), Siberian Branch of the Russian Academy of Medical Sciences is a relational database, i.e. it is organized in the form of related tables, where rows correspond to records, and columns of the tables are the attributes of relations. Wide spreading of the relational data model is primarily due to ease of presentation and the formation of the database, universality and usability of processing. [3, 5].

Database of the Biobank of RIMG consists of four main interrelated modules: 1) basic information; (2) biological samples; (3) the results of molecular genetic analysis; (4) user database (Figure 1). **Basic information** includes a description of personal data about the subject of research, health status, living and working conditions; the results of laboratory examinations; the relationship with other individuals in the database of the Biobank of RIMG and additional reference tables.

Samples of the biological material - contains information about the DNA samples, blood, cells, tissues, etc. The module includes information about the place and date of sampling of biological material, the method of DNA extraction, storage conditions and location of the sample. Also in this section the using of the samples is recorded.

The results of molecular genetic analysis represent reference table of names and descriptions of the investigated genes; a reference sequence in the SNP database (www.ncbi.nlm.nih.gov); variants of genotypes; type of polymorphism (SNP, VNTP, del, ins, etc.); photos of gels and the genome sequences in the text format. The results are presented as

summary tables, which combine following units: "individual", " biological sample", "variants of genotypes".

Users This module stores information about users of the database with an indication of the level of access, history of queries, transaction history (entering, editing, deleting data, etc.) and the list of allowed IP addresses.

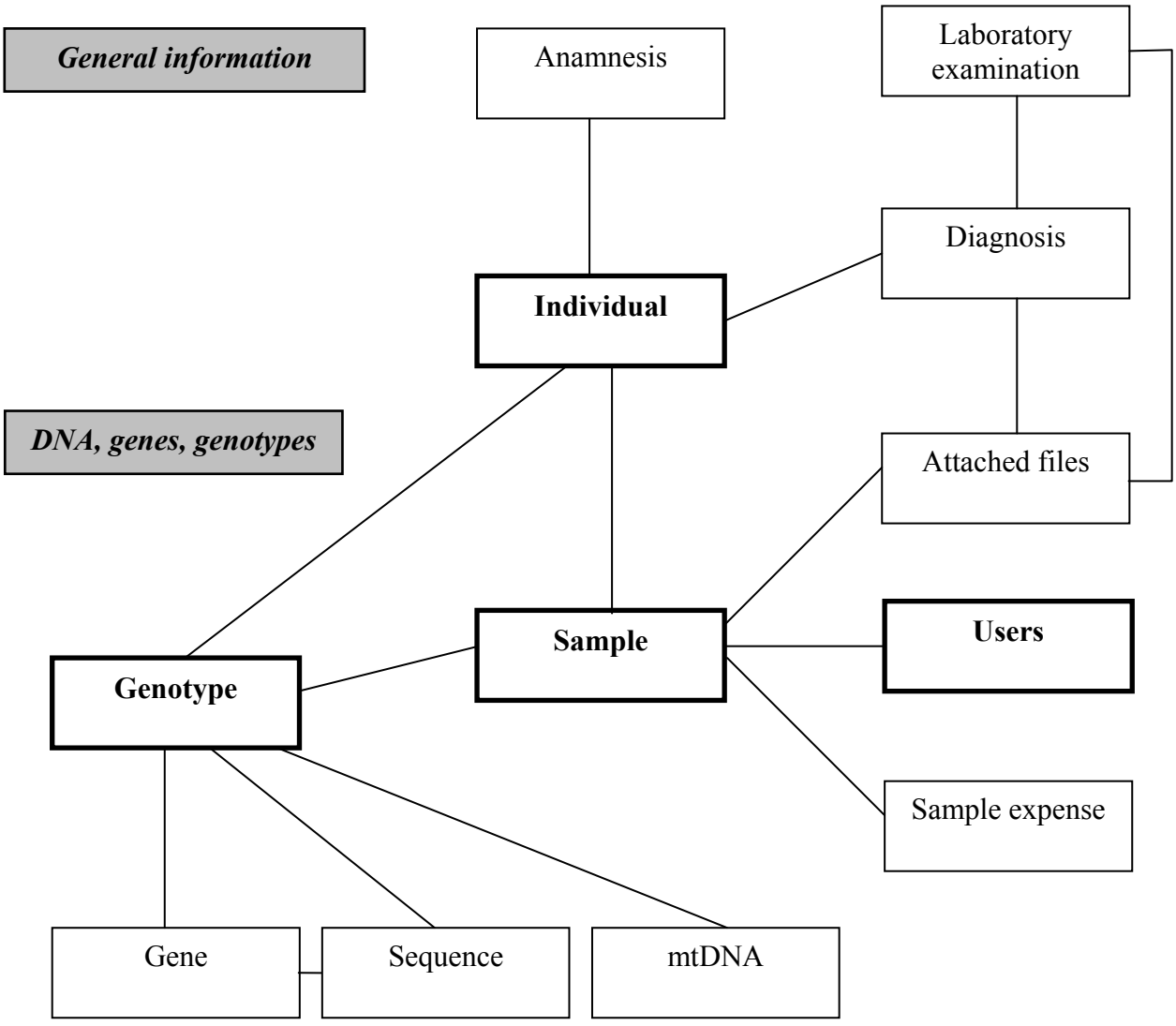
For the designing of the DB of Biobank of Institute of medical genetics of Russian Academy of Medical Science, requirements to the interface are developed which reflect specifics of the Biobank function. The interface of a DB is organized in the form of two basic levels - User and Administrator. The Administrator level is necessary for entering and editing the information in the DB by the users having corresponding access rights, and it contains the basic (entering of data into a DB) and auxiliary (editing of reference tables) interface. The administrator has rights to enter, edit and remove data in all mainframes, to set access rights to other users, and he has also an opportunity of creation of a backup copy of the DB. The User level is necessary for access to available data, formation of sample sets according the set parameters, display of results of import or export of data. It contains: The form of authorization, the form of query creation, the form of the results presentation, the form of last user queries, the form of data entering/editing, the form of viewing of user activity. To date, the designing stage for the DB of Biobank of Institute of medical genetics of Russian Academy of Medical Science is completed. For good data safety at a level of information system, it is planned to provide authorized access to the DB, restriction of access for users (only with specified IP addresses), protection against automatic selection of the password (with blocking of attacking IP addresses), access to personal data, restricted access to entering and editing of data (limited by several persons). Also storage of history of user authorization, user queries and editing history is stipulated. For the DB of Biobank of Institute of medical genetics of Russian Academy of Medical Science, operating in a local computer network is planned, with an opportunity of the protected access from external network.

Thus, in the Biobank of Institute of medical genetics of Russian Academy of Medical Science, the structure of a unique, functional, specialized electronic resource (DB) is developed which meets the requirements of experts working in various areas of medical genetics and corresponding requirements of the federal law on personal data safety [6]. This DB can be used in the medical and biological research centers, for maintenance of safety of biological collections and to increase of efficiency of scientific research.

The study is supported by Federal Target Program “Kadry” on 2009 - 2013 (The contract no. P722 from 12.08.2009).

Literature

1. Bragina E.Y., Buikin S.V. Influence of control sample size on significance of associations of genetic markers with development of multifactorial diseases // Jakutian Medical Journal. – 2009. – №2(26). – P. 142-144.
2. Bragina E.Y., Buikin S.V., Puzyrev V.P. Biobanks: problems and perspectives of their use in genetic studies of human common diseases // Medical Genetics. – 2009. – №3. – P. 20-27.
3. Ingasimutu S. Principles of bioinformatics. - Izhevsk: SRC «Regular and chaotic dynamics», Institute of computer research, 2007. - 320 p.
4. Karpova I.P. Introduction to databases. The manual. M: Publishing house MGIEM, 2005. - 96 p.
5. Kusnetsova S.D. Basics of modern databases, information-analytical materials. 2008. http://citforum.ru/database/osbd/contents.shtml
6. The Federal law 7/27/2006 N 152-FZ (redaction by 12/27/2009) " About personal data " // The Russian newspaper. 2006. - №165.
7. Feero W.G., Guttmacher A.E. Genomewide Association Studies and assessment of the risk of disease // N Engl J Med. – 2010. – V. 363. – P. 166 -176.
8. Frazier L., Sparks E., Sanner J.E. Biobanks and biomarker research in cardiovascular disease // J Cardiovasc Nurs. – 2008. – V. 23(2). – P. 153-158.
9. Gottweis H., Zatloukal K. Biobank governance: trends and perspectives // Pathobiology. – 2007. – V.74(4). – P. 206-211.
10. Little J., Higgins J.P.T., Ioannidis J.P.A. Strengthening the reporting of genetic association studies (STREGA): an extension of the STROBE Statement // Hum.Genet. – 2009. – V. 125. – P. 131-151.
11. Shickle D. The consent problem within DNA biobanks // Stud Hist Philos Biol Biomed Sci. – 2006. – V.37(3). – P. 503-519.
12. Swede H., Stone C.L., Norwood A.R. National population-based biobanks for genetic research // Genetics in Medicine. – 2007. – V.9(3). – P. 141-149.
13. The Wellcome trust Case Control Consortium. Genome-wide association study of 14000 cases of seven common diseases and 3000 shared controls // Nature. - 2007. – V. 447. – P. 661-678.



Pic.1 Structure of organization DB of Biobank of institute of medical genetics of Russian Academy of Medical Sciences