

В. В. Кириченко, М. В. Спринджук, Р. С. Сергеев, О. М. Залуцкая,
А. М. Скрягина, А. Е. Скрягин

БЕЛОРУССКИЙ ПОРТАЛ О ТУБЕРКУЛЕЗЕ: УНИКАЛЬНАЯ БАЗА ДАННЫХ В ОТКРЫТОМ ДОСТУПЕ

Белорусский портал о туберкулезе создавался в рамках программы Американского фонда гражданских исследований и развития (Civilian Research and Development Foundation, CRDF). Участниками проекта являются:

- Национальный институт аллергических и инфекционных заболеваний США (National Institute of Allergy and Infectious Diseases, NIAID);*
- Институт проблем информатики Национальной академии наук Беларуси;*
- Республиканский научно-практический центр пульмонологии и фтизиатрии.*

Уникальность Белорусского портала о туберкулезе заключается в представленной информации и данных, которые содержат не только статьи, исследования и полезную информацию о туберкулезе, но и обширную коллекцию компьютерных томографий и рентгеновских снимков легких пациентов, больных туберкулезом, содержащий полный набор клинической информации о ходе лечения заболевания. Портал предлагает пользователю гибкий и удобный инструмент поиска интересующих его случаев заболевания туберкулезом, несколько вариантов загрузки данных и удобный интерфейс.

Ключевые слова: базы данных изображений, визуализация данных, обработка изображений, туберкулез.

V. V. Kirichenko, M. V. Sprindzuk, R. S. Sergeev, O. M. Zalutskaya,
A. M. Skryagina, A. E. Skryagin

THE BELARUSIAN TUBERCULOSIS PORTAL: A NOVEL INTEGRATED RESOURCE TO FACILITATE PULMONARY MEDICINE RESEARCH

Multi-drug resistant tuberculosis is a difficult to cure and an extremely dangerous disease. Although it is seldom in Europe, collaboration is required to fight this disease worldwide. Imaging methods are the major approaches to the primary visualization and diagnosis of respiratory pathology. To enrich the scientific and educational basis of tuberculosis research, we have designed the BTB Portal (Belarusian Tuberculosis Portal, <http://www.tuberculosis.by>) containing a collection of computed tomography images of multi-drug resistant tuberculosis patients and appropriate clinical data. Web portal combines several options allowing searching tuberculosis clinical cases of interest and provides functionalities for downloading images and data. Besides the image archive, the portal contains additional information including recently published research articles and various hyperlinks to the tuberculosis related sites. The BTB Portal provides a highly functional multi-faceted resource that will aid future research and education of the modern tuberculosis.

Key words: databases of images, visualization of data, processing of images, tuberculosis.

Предпосылки

Туберкулез является основной проблемой здравоохранения Беларуси [9]. Недавно ситуация была осложнена появлением и развитием туберкулеза с множественной лекарственной устойчивостью (МЛУ), туберкулеза с широкой лекарственной устойчивостью (ШЛУ) и туберкулеза с ВИЧ (вирусом иммунодефицита человека), которые требуют длительного лечения. Самые тяжелые случаи появляются во всей стране. В 2009 году в Беларуси было зарегистрировано 4441 случай заболевания туберкулезом, из которых 22,4% являлись случаями туберкулеза с множественной лекарственной устойчивостью [7, 8, 10].

В течение последних нескольких лет в Беларуси осуществляется целый ряд проектов в области компьютеризированной диагностики болезней легких и их лечения. В частности, была разработана автоматизированная телемедицинская система по скринингу флюорографических исследований в городе Минске. Она охватывает около 20% населения страны. В результате нескольких первых лет эксплуатации, была накоплена большая база данных рентгеновских изображений. Кроме того, была разработана экспериментальная система на основе грид-сети, позволяющая ускорить анализ объемных компьютерных томографий и рентгеновских снимков для обнаружения границы опухоли легких и т.д. Эти проекты создали прочную базу для исследования проблемы туберкулеза, которое будет осуществляться специалистами из ведущих медицинских учреждений Беларуси.

В мире также существует ряд веб-сайтов, предназначенных для изучения туберкулеза: интегрированная база данных, обеспечивающая доступ к данным генома туберкулеза и ресурсов, имеющих отношение к открытию и разработке противотуберкулезных препаратов и вакцин (www.tbdb.org); ресурс о мутации лекарственной устойчивости микобактерий туберкулеза (www.tbdreamdb.com); ресурс, помогающий найти информацию о туберкулезе по всему миру (www.findtbresources.org); портал центра по контролю и профилактике заболеваний, отдела ликвидации туберкулеза (www.cdc.gov/tb); портал Международного противотуберкулезного центра Карри (www.nationaltbcenter.ucsf.edu); портал Национального туберкулезного центра Хартленд (www.heartlandntbc.org); портал Юго-Восточного Национального противотуберкулезного центра (<http://sntc.medicine.ufl.edu>); веб-сайт о туберкулезе в Канаде (www.hc-sc.gc.ca/hc-ps/dc-ma/tuberculosis-eng.php) [1–6]. На всех перечисленных выше ресурсах можно найти исчерпывающую информацию о туберкулезе, однако там практически нет материалов для изучения, таких как файлы компьютерной томографии, рентгеновские снимки больных туберкулезом пациентов. Именно наличие таких материалов делает Белорусский портал о туберкулезе (www.tuberculosis.by) уникальным ресурсом, позволяющим исследователям получить большой набор данных для исследования. Пользователи Белорусского портала о туберкулезе имеют доступ к самому большому набору реальных случаев заболеваний туберкулезом, имеют возможность ознакомиться с клиническими данными, изучить файлы с компьютерной томографией. Так же Белорусский портал

о туберкулезе предоставляет набор средств, полезных для пользователя, таких как удобный многопараметрический поиск, несколько вариантов загрузки данных, большой архив компьютерных томографий пациентов, больных туберкулезом, включающий 9280 случаев, а также набор из 401 случай заболевания туберкулезом, включающий в себя как медицинские изображения, так и клиническое описание каждого случая (результаты лабораторных исследований, рентгенологический протокол, комментарии врачей).

Материалы и методы

Программные компоненты

При разработке Белорусского портала о туберкулезе и программного обеспечения, предназначенного для наполнения данных портала, были использованы следующие технологии и программные платформы:

- Microsoft .NET Framework 4.5 (www.microsoft.com);
- Asp.NET MVC 4.0 (www.asp.net/mvc);
- ADO. NET ([http://msdn.microsoft.com/en-us/library/e80y5yhx\(v=vs.110\).aspx](http://msdn.microsoft.com/en-us/library/e80y5yhx(v=vs.110).aspx));
- набор библиотек ClearCanvas (www.clearcanvas.ca);
- шаблоны, основанные на Bootstrap (<http://getbootstrap.com>);
- JavaScript библиотека jQuery (<http://jquery.com>).

База данных

В процессе разработки портала и анализа данных, было принято решение, что наиболее подходящим и функциональным вариантом будет вариант хранения данных в формате JSON (JavaScript Object Notation), а для поиска данных использовать библиотеку Lucene.NET. Такое сочетание дает широкие возможности масштабируемости данных без необходимости перерабатывать структуру базы данных.

Одна запись содержит преобразованные данные для мгновенного отображения пользователям. Т.е. входные данные обрабатываются и преобразуются таким образом, что JSON запись находится в читабельном представлении, например: «SocialData»: «lived in private house with 1 persons, alcoholic, smoked at least for the last 5 years, education «complete school (a-level, gymnasium)».

Источники данных:

- Данные для портала аккумулируются из четырех основных источника данных:
- база данных Access
- Excel файл с результатами лабораторных анализов на лекарственную чувствительность
- Excel файл с результатами лабораторных исследований мокроты
- директория в файловой системе с DICOM файлами с результатами компьютерной томографии и рентгеновскими снимками.

Базу данных Access разработали и наполняют данными на базе Республиканского научно-практического центра (РНПЦ) пульмонологии и фтизиатрии. В данную базу данных попадают пациенты, вовлеченные в проект. Основным критерием для попадания пациента в базу данных является наличие у него компьютерной то-

мографии и диагностированного заболевания туберкулезом.

Excel файл с результатами лабораторных анализов на лекарственную чувствительность и Excel файл с результатами лабораторных исследований мокроты формируют в лаборатории РНПЦ пульмонологии и фтизиатрии на основе проведенных исследований.

Директорию с DICOM файлами с результатами компьютерной томографии и рентгеновскими снимками формируют специалисты РНПЦ пульмонологии и фтизиатрии. Данный процесс является ручным, в рамках которого специалист извлекает из клинической базы данных все исследования для конкретного пациента и помещает в созданную директорию.

Процедура сбора данных

Для автоматизации поступления данных на портал, было разработано программное обеспечение, которое выполняет следующие функции:

- анонимизация данных пациента
- валидация данных
- преобразование данных в CSV (Comma-Separated Values) файл, каждая запись которого имеет следующую структуру: clinic_record_id, date (yyyy-MM-dd), event_type, field_name, field_value
- отправка данных на FTP (File Transfer Protocol) сервер портала.

Схема работы программного обеспечения заключается в следующем. Пользователь (в нашем случае, это специалист РНПЦ пульмонологии и фтизиатрии) предоставляет один из возможных источников данных, программное обеспечение конвертирует предоставленные данные в файл CSV и отправляет его на FTP сервер. На FTP сервере реализован сервис, который через определенный интервал времени анализирует содержимое папки и в случае, если данные обновились, происходит автоматическое чтение и преобразование данных из формата CSV в JSON, а также генерация индексов Lucene.NET, которые используются в поисковой системе.

Результаты и обсуждение

Интерфейс пользователя

Структура портала

Белорусский портал о туберкулезе представляет собой уникальный ресурс, который содержит статьи, исследования и полезную информацию о туберкулезе, коллекцию компьютерных томографий и рентгеновских снимков легких пациентов, больных туберкулезом. Условно портал можно разделить на 2 составляющих – информационная и практическая.

Информационная составляющая представляет собой следующие разделы:

- раздел с информацией о состоянии проблемы туберкулеза в Беларуси;
- раздел со статьями о туберкулезе, соавторами которых являлись участники проекта;
- раздел с подборкой полезных ссылок на международные ресурсы о туберкулезе;

– раздел с результатами исследовательской работы в рамках проекта.

Практическая составляющая состоит из двух основных частей:

- архив КТ исследований без детальной клинической информации;
- КТ исследования с детальным клиническим описанием.

Раздел «CT Images with clinical data» представляет собой каталог компьютерных томографий с детальной клинической информацией. Этот каталог является уникальным набором случаев заболевания туберкулезом. На момент написания статьи, в разделе содержится 401 случай. Каждый случай содержит компьютерную томографию в стандарте DICOM, социальные данные о пациенте (например, рос, вес, возраст, условия проживания, признак алкоголизма и пр.), историю лечения (промежуточные исследования, рентген-снимки, тесты на лекарственную чувствительность и пр.), диагноз и результат лечения. Большинство случаев в этом разделе являются случаями туберкулеза с широкой лекарственной устойчивостью.

В этом разделе реализована гибкая поисковая система. Пользователь может сам формировать запросы, написав их в соответствующем поле (рис. 1, метка 1), а также может воспользоваться функцией «Query Builder» (рис. 1, метка 3). В разделе «Query builder» перечислены все возможные параметры, которые обрабатывает поисковая система. Напротив параметра, в скобках, указаны возможные действия с параметром, например

By age (":", ">", "<", ">=", "<="), и возможные значения, например, drug_resistance: «mdr non xdr». Так же после названия параметра приведен пример его использования, кликнув на который, выражение попадет в поле формирования запроса (рис. 1, метка 1).

После выполнения запроса, пользователь может увидеть сводную информацию по результатам запроса (рис. 1, метка 2), в которой перечислено количество случаев туберкулеза в разрезе пола пациента, возраста и вида лекарственной устойчивости туберкулеза.

Информация о случае представлена отдельным блоком (рис. 1, метка 4), в котором все данные разделены на 2 блока: общая информация о случае (Problem) и история лечения (History). В блоке Problem перечислены социальные и анкетные данные пациента, комментарии врача. В блоке History перечислены события, происходившие во время лечения, например, тесты на лекарственную чувствительность, компьютерные томографии и т.п.

Пользователю портала предоставляется несколько вариантов загрузки данных:

– Возможность загрузить конкретный случая целиком, включая DICOM файлы и клиническое описание. Сделать это можно нажав кнопку «Download» (рис. 1, метка 5) напротив ID случая.

– Возможность загрузить только клинические данные в формате csv. Сделать это можно нажав кнопку «Download Clinical Records (No Images)» (рис. 1, метка 6). При этом пользователю будет предоставлен для загрузки ZIP архив, включающий в себя CSV файл с описа-

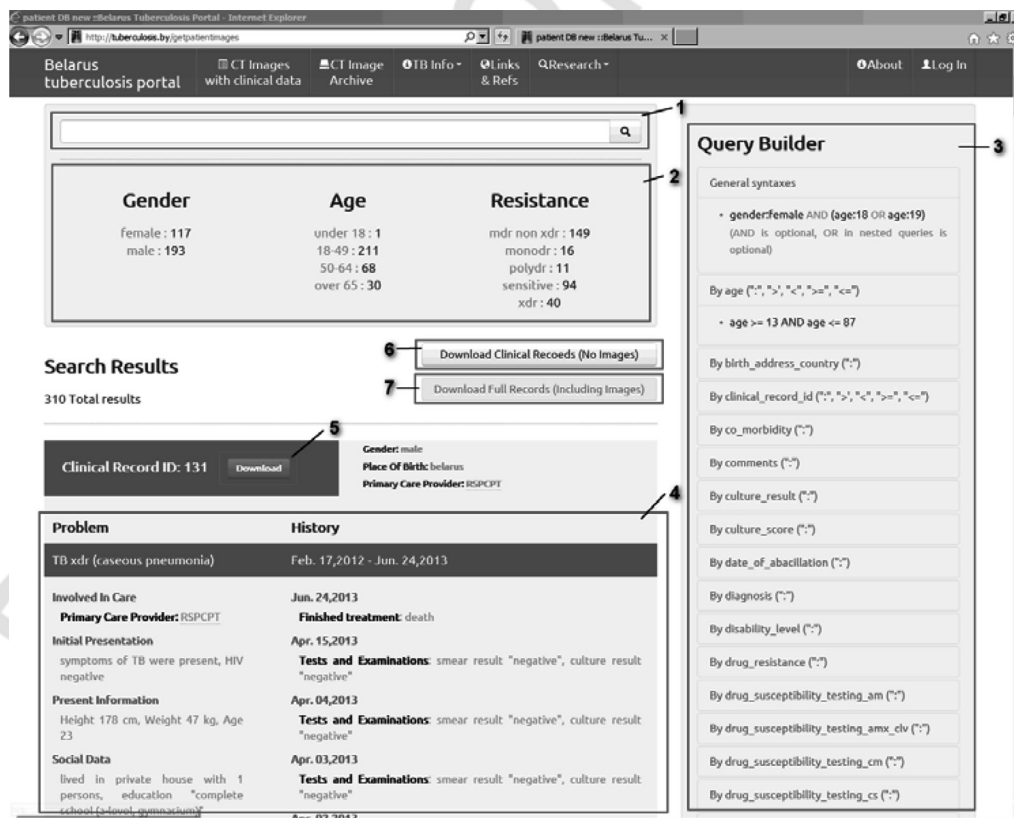


Рис. 1. CT Images with clinical data

нием случаев, попавших в запрос и текстовый документ с текстом запроса.

– Возможность загрузить полный набор данных, включающий клиническое описание и файлы DICOM. Сделать это можно нажав кнопку «Download Full Records (Including Images)» (рис. 1, метка 7). При этом пользователю будет предоставлен для загрузки ZIP архив, включающий в себя CSV файл с описанием случаев, попавших в запрос, текстовый документ с текстом запроса и папку с DICOM файлами компьютерной томографии или рентгеновских снимков. Важное замечание: кнопка «Download Full Records (Including Images)» будет не активна до тех пор, пока в результате запроса больше 30 случаев.

Раздел «CT Image Archive» представляет собой архив компьютерных томографий, который включает 9280 исследований пациентов, больных туберкулезом. Каждый случай включает в себя цифровое изображение в стандарте DICOM. DICOM файлы содержат в себе только информацию об исследовании и цифровое изображение.

В данном разделе реализована удобная система отбора исследований, которая позволяет сузить набор исследований, указав пол и возраст пациента (рис. 2, метка 1). В результате отбора будут отображены исследования, попавшие под критерии отбора. Каждый случай отображен в виде фронтальной проекции и набора случайных срезов из DICOM файла компьютерной томографии (рис. 2, метка 2). Основное исследование, на основе которого был поставлен диагноз, отображается большим изображением, в левом верхнем

углу которого указана дата проведения исследования, возраст пациента на момент исследования и количество серий. В случае, если у данного пациента есть более одной компьютерной томографии, они отображаются в виде более маленьких изображений рядом с основным.

После каждого отбора, пользователю выводится информация о том, сколько всего случаев попало в отбор и сколько отображено в данный момент (рис. 2, метка 3).

Пользователь имеет возможность загрузить любой случай, но эта возможность доступна только зарегистрированным пользователям.

Архив файлов компьютерной томографии представляет большую исследовательскую ценность и может быть полезна при изучении проблемы туберкулеза, научных исследованиях и в обучающих целях.

В разделе «TB Info» можно найти статьи о состоянии туберкулеза в Беларуси, существенные факты о туберкулезе. В подразделе «Articles on tuberculosis» находится набор статей, опубликованных в международных журналах, и соавторами которых являются участники проекта.

В разделе «Links & Refs» собраны полезные ссылки на международные порталы о туберкулезе.

В разделе «Research» опубликованы результаты теоретического исследования возможности раннего выявления случаев туберкулеза с широкой или множественной лекарственной устойчивостью.

В разделе «About» можно увидеть информацию о портале, проекте, целях проекта и участниках проекта.

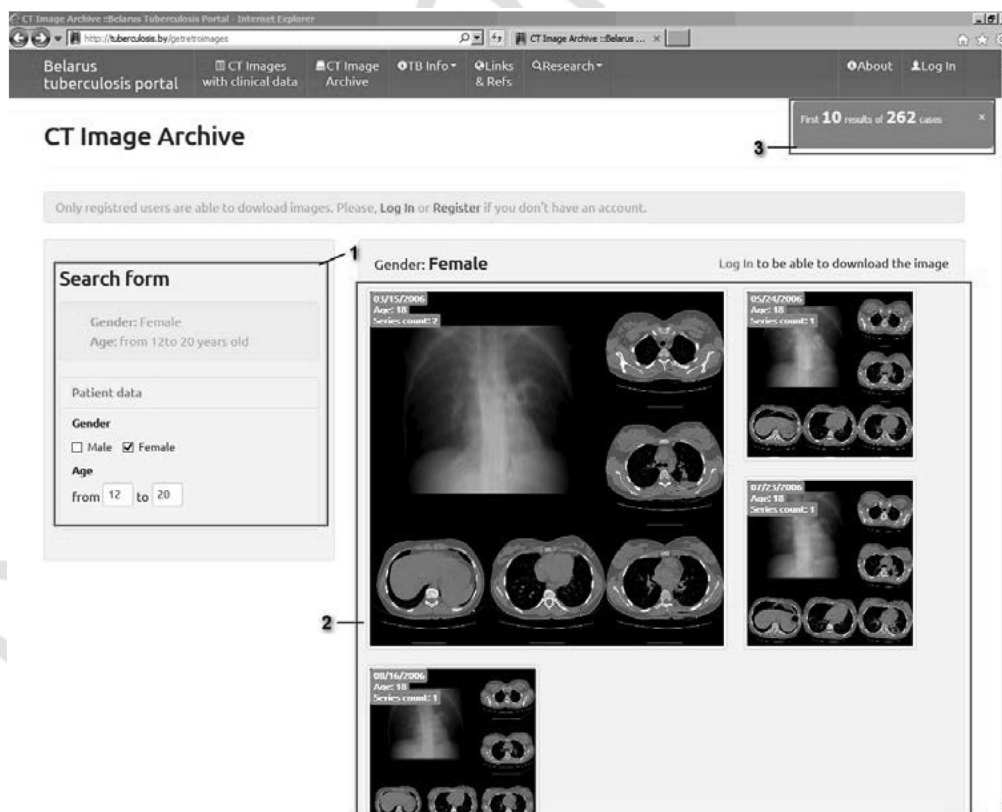


Рис. 2. CT Image Archive

Дальнейшая разработка Белорусского портала о туберкулезе

В ближайшем будущем планируется существенно расширить как функциональность портала, так и набор данных. Так, будет реализована возможность пользователю сохранять поисковые запросы для дальнейшего использования, вести их историю, что даст возможность исследователям обмениваться запросами для совместного обсуждения особенно интересных случаев. Ввести ссылки на GenBank (<https://www.ncbi.nlm.nih.gov/genbank>) и PATRIC (Pathosystems Resource Integration Center, <http://patricbrc.org/portal/portal/patric/Home>) для тех случаев туберкулеза, в которых проводилось секвенирование ДНК микобактерий туберкулеза. Данные расширятся информацией о ВИЧ, результатами проводившихся тестов, таких как Hine-тест, Вастес и GeneExpert. Так же будет увеличиваться количество случаев туберкулеза с полным клиническим описанием.

Нами разработан ресурс, доступный всем желающим, который предоставляет большой набор уникальных данных. Эти данные могут использоваться в обучающих целях, т.к. на Белорусском портале о туберкулезе есть большой набор DICOM файлов с компьютерной томографией и рентгеновскими исследованиями туберкулеза. Исследователи могут найти на портале много информации о туберкулезе в Беларуси и мире, проводить исследования на интересующих его когортах пациентов, изучать медицинские данные случаев заболевания туберкулеза.

Мы сфокусировали наши усилия на разработке ресурса, который будет полезен и интересен широкому кругу пользователей, постарались наполнить его уникальными данными, которые невозможно найти ни на одном другом публичном ресурсе. Реализовали очень гибкую и удобную систему поиска, постоянно пополняем базу данных новыми случаями и расширяем набор данных новыми параметрами.

В настоящее время на портале собран большой архив компьютерных томографий, включающий 9280 случаев и архив полных историй болезни пациентов, больных туберкулезом, включающий в себя полный набор клинических данных, результаты лабораторных исследований, компьютерная томография и рентгеновские снимки. В этом архиве на сегодня находится 401 случай.

Благодарности

Мы выражаем благодарности Американскому фонду гражданских исследований и развития (CRDF), в рамках программы которого был разработан Белорусский портал о туберкулезе, а также Национальному институту аллергических и инфекционных заболеваний США (NIAID) и Республиканскому научно-практическому центру пульмонологии и фтизиатрии (Беларусь) за содействие и помощь в накоплении данных, а также сообществам разработчиков Microsoft, jQuery, CrearCanvas и Bootstrap, чьи продукты использовались при разработке.

Литература

1. *Bharti, R., Das R., Sharma P., et al.* MTCID: a database of genetic polymorphisms in clinical isolates of Mycobacterium tuberculosis. *Tuberculosis (Edinb)* // 2012. № 2. P. 166–172.
2. *Hassan, S., Logambiga P., Raman A. M., et al.* MtbSD – a comprehensive structural database for Mycobacterium tuberculosis. *Tuberculosis (Edinb)* // 2011. № 6. P. 556–562.
3. *Jena, L., Wankhade G., Kumar S., et al.* MTB-PCDB: Mycobacterium tuberculosis proteome comparison database. *Bioinformation* // 2011. № 3. P. 131–133.
4. *Lew, J. M., Mao C., Shukla M., et al.* Database resources for the tuberculosis community. *Tuberculosis (Edinb)* // 2013. № 1. P. 12–17.
5. *Reddy, T. B., Riley R., Wymore F., et al.* TB database: an integrated platform for tuberculosis research. *Nucleic Acids Res* // 2009. № 1. P. D499–508.
6. *Sandgren, A., Strong M., Muthukrishnan P., et al.* Tuberculosis drug resistance mutation database. *PLoS Med* // 2009. № 2. P. e2.
7. *Skrahina, A., Hurevich H., Zalutskaya A., et al.* Multidrug-resistant tuberculosis in Belarus: the size of the problem and associated risk factors. *Bull World Health Organ* // 2013. № 1. P. 36–45.
8. *Skrahina, A., Hurevich H., Zalutskaya A., et al.* Alarming levels of drug-resistant tuberculosis in Belarus: results of a survey in Minsk. *Eur Respir J* // 2012. № 6. P. 1425–1431.
9. *Titov, L. P., Hancharou A. Y., Skrahin A. Y., et al.* Immunotherapy of patients with multiple and extreme drug-resistant pulmonary tuberculosis with autologous monocyte- and stem cell-derived dendritic cells vaccine. *International Journal of Mycobacteriology* // 2015. № 3. P. 58.
10. *Titov, L. P., Zakerbostanabad S., Slizen V., et al.* Molecular characterization of rpoB gene mutations in rifampicine-resistant Mycobacterium tuberculosis isolates from tuberculosis patients in Belarus. *Biotechnol J* // 2006. № 12. P. 1447–1452.