

The RD5000 Database: Facilitating Clinical, Genetic, and Therapeutic Studies on Inherited Retinal Diseases

Ramon A. C. van Huet,¹ Clasien J. Oomen,^{1,2} Astrid S. Plomp,³ Maria M. van Genderen,⁴ B. Jeroen Klevering,¹ Reinier O. Schlingemann,⁵ Caroline C. W. Klaver,^{6,7} L. Ingeborgh van den Born,⁸ Frans P. M. Cremers^{2,9}; and the RD5000 Study Group

¹Department of Ophthalmology, Radboud University Medical Center, Nijmegen, The Netherlands

²Department of Human Genetics, Radboud University Medical Center, Nijmegen, The Netherlands

³Department of Clinical Genetics, Academic Medical Center, Amsterdam, The Netherlands

⁴Bartiméus, Institute for the Visually Impaired, Zeist, The Netherlands

⁵Department of Ophthalmology, Academic Medical Center, Amsterdam, The Netherlands

⁶Department of Ophthalmology, Erasmus Medical Center, Rotterdam, The Netherlands

⁷Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands

⁸The Rotterdam Eye Hospital, Rotterdam, The Netherlands

⁹Radboud Institute for Molecular Life Sciences, Radboud University Medical Center, Nijmegen, The Netherlands

Correspondence: Frans P. M. Cremers, Department of Human Genetics (855), Radboud University Medical Center, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands; Frans.Cremers@radboudumc.nl

See the Appendix for the members of the RD5000 Study Group.

Submitted: July 24, 2014

Accepted: July 25, 2014

Citation: van Huet RAC, Oomen CJ, Plomp AS, et al. The RD5000 database: facilitating clinical, genetic, and therapeutic studies on inherited retinal diseases. *Invest Ophthalmol Vis Sci*. 2014;55:7355–7360. DOI:10.1167/iovs.14-15317

Inherited retinal diseases (IRDs) represent a clinical and genetic heterogeneous group of chorioretinal disorders. The frequency of persons affected by an IRD due to mutations in the same gene varies from 1 in 10,000 to less than 1 in a million. To perform meaningful genotype-phenotype analyses for rare genetic conditions, it is necessary to collect data from sizable populations. Although several standardized functional tests are used widely, ophthalmologic data usually are stored in local databases and not in multicenter databases that are linked with other centers. To be able to register ophthalmologic data of all Dutch patients with IRDs into one database, we developed the RD5000 database (RD5000db), which can harbor all ophthalmologic and selected genetic data. Authorization rights for the management, data entry, and data sharing have been set up, rendering this database into a user-friendly, secure, and widely used repository that will facilitate future studies into molecular genetics and therapies for IRDs. The RD5000db database has the potential to grow into a European standard for the registration of data from IRDs.

Keywords: retinal disease, registry, genetic diseases

Inherited retinal diseases (IRDs) constitute a clinically heterogeneous group of disorders that involve degeneration of the photoreceptors, RPE, and/or choroid.^{1–4} The incidence rates vary from 1 in 10,000 (for Stargardt disease) to less than 1 in a million for rare clinical subtypes.^{5–9} The diseases affect both sexes, and include major causes of blindness in children and young adults worldwide, although some patients remain asymptomatic until middle-age. The phenotypic variability is caused by genetic defects in a large number of genes. In the last 25 years, over 200 genes associated with syndromic and nonsyndromic IRDs (RetNet, available in the public domain at <https://sph.uth.edu/retnet/>) were identified.^{4,10,11}

The internet offers several public and license-protected databases that contain genetic data on IRDs, such as the Leiden Open (source) Variation Databases (LOVDs, available in the public domain at <http://www.lovd.nl/3.0/home>), the Human Gene Mutation Database (HGMD, available in the public domain at <http://www.hgmd.org/>), and the retina international mutation database (available in the public domain at <http://www.retina-international.org/sci-news/databases/mutation-database/>). Still, databases containing clinical data are, to our knowledge, not publically available. Instead, most specialized medical centers use a database exclusively accessi-

ble by that specific center. This lack of publically available clinical data most likely is due to privacy legislation and medical ethics. On the other hand, there are websites that provide an overview of genetic and clinical features of IRDs published in the literature, such as the Retinal Network (RetNet, available in the public domain at <https://sph.uth.edu/retnet/>) and the Online Mendelian Inheritance in Man (OMIM, available in the public domain at <http://www.omim.org/>).

Since incidence rates are low in IRDs, national collaborations are the first step to get an overview of cohort sizes, the distribution of certain genetic defects, and the matching clinical features. This also will be important to identify persons for (gene) specific therapeutic trials. In this light, 11 Dutch ophthalmic centers collaborated in a project, named RD5000, to set up a database with anonymized clinical and genetic data: the RD5000 database (RD5000db, available in the public domain for members at <http://www.RD5000db.nl>). The name “RD5000” was based on the estimated prevalence of 5000 patients with inherited retinal diseases in the Netherlands. To obtain larger cohort sizes for studies on specific genotypes, phenotypes, or genotype-phenotype correlations, international expansion of the database is possible and desirable.

OBJECTIVES OF THE RD5000 DATABASE

To our knowledge, the RD5000db is the first multicentered, web-based database available in the field of inherited retinal diseases. Main objectives are standardization of clinical and genetic data registration, pseudo-anonymized (“pseudomized”) storage of data, and web-based, controlled data sharing among participating medical centers. The pseudomized character of the data storage lies in the absence of personal characteristics that could lead to the identification of a specific patient, and the presence of a center-specific internal patient identification number. This results in anonymized data for the centers that have access to the database, and personalized data for the patient’s medical center that also owns the source data files. Data Access by other centers is controlled; only a few general features are available commonly in a “catalogue.” Access to all other details can be granted either for clinical reasons (e.g., when the patient attends an ophthalmologist or geneticist in another center participating in RD5000db) or research reasons (e.g., for a specific research question; see Data Access below). This controlled access allows the professionals to be in charge of the data to be shared.

DATABASE DESIGN

We reviewed literature on existing databases in the field of ophthalmic retinal disease. The Ophthabase database (used in Tuebingen, Germany) was used for database structure and layout.¹² Additionally, the biobank of the string-of-pearls initiative (The Netherlands, Dutch: Parelnoer initiatief) was used for their way of structuring multicenter data entry and access to patient data (see Data Access).¹³ Additionally, jurisdiction and expert opinion on national privacy legislation and patient data were obtained.

Source Code

The source code of Ophthabase (developed by Institute for Ophthalmic Research, Tübingen, Germany, as part of the EVI-GENORET Project) was generously provided for further development. The HTML design and data model of Ophthabase were used as a starting point for RD5000db and were altered to the specifications of RD5000db by the company First8 (available in the public domain at www.first8.com). The RD5000db is built in Java 1.7 (Oracle Corporation, Redwood Shores, CA, USA) and runs on an Apache Tomcat 7.0 (The Apache Software Foundation, Forest Hill, MD, USA) application server, and uses PostgreSQL 8.4 (EnterpriseDB Corporation, Bedford, MA, USA) as a relational database. The web interface is built using Apache Wicket 6 (The Apache Software Foundation) and uses Spring Core (Pivotal Software, Palo Alto, CA, USA) framework to interact with services, which contain the business logic. Hibernate (Red Hat, Inc., Raleigh, NC, USA), which implements the Java Persistence application programming interface 2 standard, is used to query the database. Spring Security (Pivotal Software) is used for authorization and authentication, and is used to implement the roles and rights system. Access from the web always is redirected from http to https, which uses Secure Socket Layer (SSL) to ensure encrypted transportation of the data over the internet.

Data Entry

The RD5000db is set up as a research database that contains pseudomized patient data. The internal IDs are used to relate the data stored in RD5000db to information in the medical files of the centers in which the patient is treated (pseudomization). Although no personal data that could identify a patient are

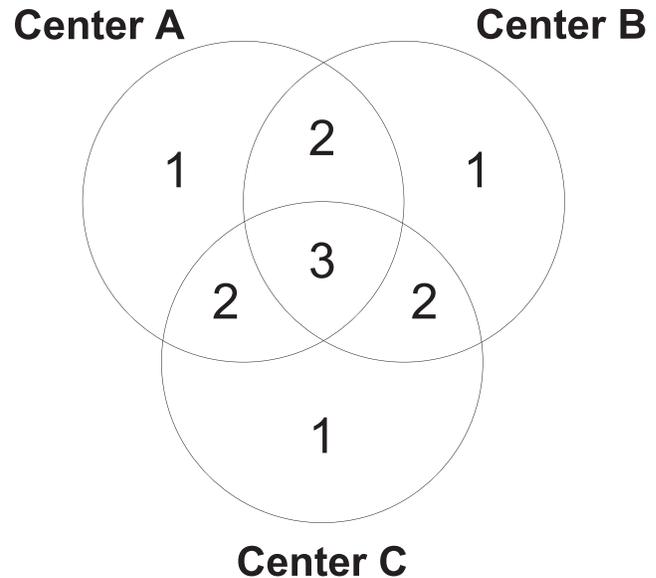


FIGURE 1. Visualization of data access in case patient data are entered from one, two, or three centers. The *numbers* indicate the number of centers that have access to the patient’s data. If more than three centers are involved, the principle is followed analogously.

stored in the database, initials, date of birth, and sex are entered to construct a hash code that is used to identify double entries. Since only the hash code is stored, the initials and date of birth are not accessible, but the year of birth and sex remain visible to users.

The data forms are organized around the pseudomized Basic Patient Data (Supplementary Fig. S1B). The structure of RD5000db was based on the case report form (CRF) and standard operative procedures (SOPs) that were formulated by the RD5000 consortium, as well as on the data structure of Ophthabase. We designed five different data forms: patient history, genetics/DNA sample, clinical examinations, technical examinations, and diagnoses forms. More details on the different data entry forms are provided in the Supplementary Appendix S1 and Supplementary Figures S1A through S1G. Numerous clinical and technical examinations can be entered when follow-up data (longitudinal data) are available. Furthermore, image files of imaging techniques or examination reports can be uploaded to support the clinical data. The database allows patients to be registered in multiple centers that all can contribute to the data volume of a single the patient (Fig. 1), which is not uncommon in The Netherlands. Family data are entered initially in the family history section of the history form (see Supplementary Appendix S1). These data only include the number of affected individuals within the family, their relation to the patient and (if available) internal registration numbers of the parents. However, RD5000db provides the ability to assign family members that are entered in the database to a family ID in the Family tab (Supplementary Fig. S2).

Informed Consent

We endeavored to enroll patients in the RD5000db with their approval to adhere to the tenets of the Declaration of Helsinki, although the data entered in the database are pseudomized and patient’s personal details are available only to the center the patient attends (via the internal ID), which makes informed consent not obligatory. Nevertheless, we created an informed consent to obtain patient approval for data entry in the database.

Privacy

Due to the sensitive nature of the collected data we carefully considered how to handle privacy correctly according to the Dutch privacy regulations. At first the idea was to encrypt all the data using a public key infrastructure (PKI). In this set-up, only the super administrator of RD5000db would have access to the private key. However, in this approach maintenance by our hosting provider or ICT implementation party would be complicated. Therefore, the data were anonymized in such a way that it is not possible to directly identify a patient. However, the center that a patient attends can identify the patient by means of a separate registration system that is not connected in any physical or electronic way to the RD5000db application. In the RD5000db, this is possible by the internal ID that is linked to a patient in that specific center; hence, only members of that center can acquire the patient's privacy-sensitive details from their own center's registration systems. Only a very limited amount of information is available to members of the other centers with access to the RD5000db. They can, however, obtain additional anonymous patient data upon permission from the review committee (see Data Access).

Identifying Double Entries

We introduced a hash code based on date of birth, initials, and sex to prevent double entry of a patient. However, this procedure has its limitations: initials are not always completely registered in the hospital's registration systems and, in some cases, names may be noted incorrectly on the medical files. For some individuals the distinction between first and family names is not clear, resulting in a different order of the initials that constructs a different hash code. Preventing double entry, however, is important, especially in these small groups of patients. When a double entry is suspected, the system administrator executes an audit trail based on year of birth, sex, and diagnosis, and verifies personal details of the presumed double entries with the responsible physicians (who generally are from two different centers), before merging the entries.

Search Engines

The RD5000db includes two search engines to explore patient data: the search engine of the database located on the Catalogue tab that searches all entries, and the search engine on the Patients tab that searches only patients accessible to the user (Supplementary Figs. S3A, S3B). Both search engines are identical in composition, but differ in function. The Catalogue tab presents a limited set of data for all patients in the database, allowing the users to have an indication of the population that has been entered in RD5000db. This information can be used for formulating a research question and obtaining data access to other patients, the procedure of which is elaborated below (see Data Access). The search function in the Patients tab provides the full data of the patients entered by the user's center, or of the patients the user has obtained access to.

Authorization Roles

The hierarchy in authorization roles in RD5000db involves one system administrator, who has access to all data and roles of all users database-wide, and multiple center administrators, who have access to all data and roles of users in their center. Each center has two additional roles available: writers, who can read, edit, and add data to patients of their center, as well as add patients to the database, and readers, who have read-only

access to data from patients of their center. The reader and writer roles can be assigned to multiple users.

The system administrator has the main supervising role in the system, which includes organizational tasks, including add/delete roles of center administrators, handle requests to delete patients from the database, merge patient's data in case of multiple entries of a single patient, change access rights per patient per form, and search all patients (see Data Access). The center administrator has the supervising role within a center and takes care of communication to the system administrator, provides help to the users in his center, adds and removes users, and changes user roles of his center, and provides passwords to users within the center.

Data Access

The database is available in the public domain at <https://www.RD5000db.nl>. To register for the database, the registration form must be completed, including a specific center key (provided by the center administrator). After registration, a password will be sent to the user, to log into RD5000db and access patient data. By default, the user has access to all the patients linked to its own center. Which actions the user can perform depends on the role assigned to him (see Authorization Roles).

Patients linked to two or more centers can be accessed by each of these centers (Fig. 1). Regulations on data management state that one center does not alter data entered by another center. Since technical and clinical examinations are stored as separate files for different time points, additional data can be added easily. Violations are tracked in the audit trails (see Audit Logging). When a patient with initials, date of birth, and sex identical to an existing entry is entered into the database, users receive the double entry notification. After verification of the patient details, the system administrator can assign the patient to both centers.

For specific research questions, an RD5000db-user can receive access to the data of patients of other centers by formulating a research question application for the RD5000db review committee (Fig. 2). The RD5000db review committee consists of RD5000-members, one of each participating center. A positive decision from this committee enables the system administrator to grant the researcher access to the requested data (Fig. 2). The access granted can be specified up to the level of specific forms (without access to the data on other forms of this patient). This same procedure is applicable in case an external researcher would wish to have data from the RD5000db. The only difference is that in case the request is granted, data are sent in an export file to the external researcher instead of granting access to the database.

Data Export

Patient data can be exported in two forms. Firstly, each single form can be exported in PDF format and can be stored in patient's medical files if desired. Secondly, in the Reporting tab patient data can be exported to a Windows Excel 2007 Worksheet format (.xlsx; Microsoft Corporation, Redmond, WA, USA), which enables further data analyses by various statistical programs. The Excel export file includes the data of all patients to which the user has access. Any filtering of the data should be done after data export. The variables are organized in separate tabs of the worksheet based on the forms of the database. In each tab, the data are arranged based on RD5000 ID. Longitudinal data, such as clinical and technical examination data of multiple visits, are presented in a single row per examination, each identified with the RD5000 ID and the examination date. Where value lists are used in the database, numeric values are present in the export worksheet,

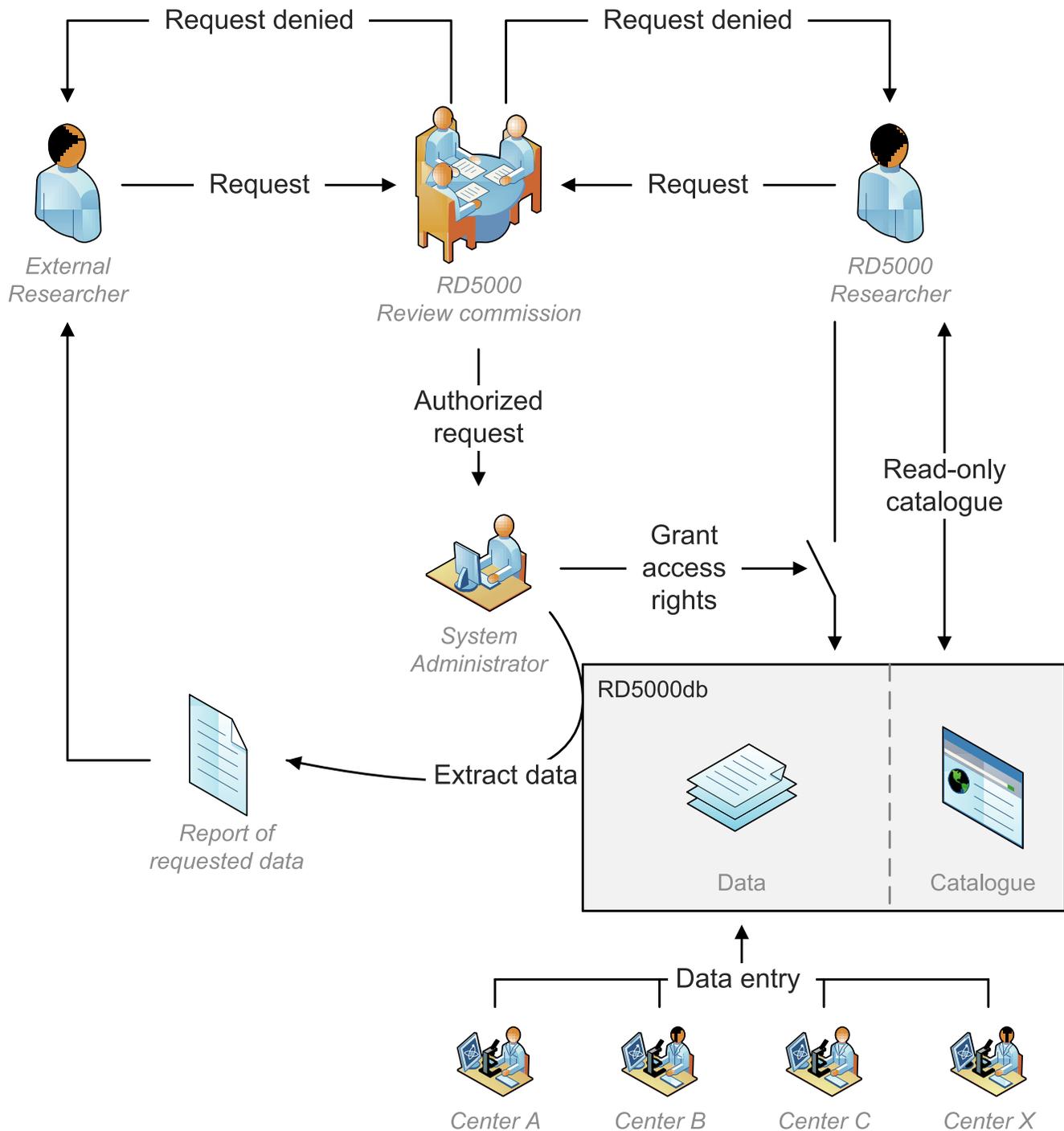


FIGURE 2. Flowchart on the process to request access to data. Each center in principle only has access to the data entered by their own center. Colleagues from other centers obtain access to the data from other centers only after approval by a review committee, which contains representatives of all data-contributing centers. External researchers can only obtain data through this review committee. Upon authorization, the system administrator will give internal researchers access to the data, and provide the requested data to the external researchers in a report file.

which can be identified using the codebook with the corresponding labels that are available in the last tab of the worksheet.

An SPSS syntax document was written to simplify data import and management in SPSS (available in the public domain for RD5000 members at <http://www.RD5000.nl>). By running the syntax, the data of all Excel worksheet tabs are imported into SPSS synthesizing one entry with correct labeling and subheadings. The syntax files are structured with

subheadings, which enables selective use of parts of the syntax code.

Audit Logging

All actions (inserts, updates and deletes) toward the database are stored in an audit log (also in the database). With this audit log, a registration is made that includes the details of the

specific user, the time of the action, as well as the details of the actions. These details can be retrieved upon request.

CURRENT STATUS AND USABILITY OF THE RD5000 DATABASE

The RD5000db was launched online at September 1, 2012. In the meantime, 1923 patients with retinal dystrophies have been registered in the database (reference date June 1, 2014), with a mean age of 41 years (range, 1-102 years). In this cohort, 62 different diagnoses were made based on clinical phenotype and inheritance pattern. Clinical and/or technical examinations were registered for 469 patients (24.3%) with an average number of visits of 4.5 (range, 1-17 for technical examinations and 1-22 for clinical examinations). Follow-up periods varied between 0 and 51 years. The DNA samples were available for 1235 patients (64.2%). Of these, 684 patients (35.6% of all entries) were solved genetically, and 188 additional patients (9.8% of all entries) carried putatively causative mutations, but have not yet been solved completely. Causative or putatively causative mutations were found in 77 different genes.

Time needed for entry of one patient is dependent mainly on the source data and the amount of details to be entered. Initially, comprehensive data entry takes up to 40 minutes per patient on average. However, for a trained person this time may be reduced to a maximum of 25 minutes for a full data record.

FUTURE PLANS

Data Quality Surveillance

In the near future, we will cope with some of the problems we encountered in the past two years. One issue of high priority is the double entry check that is intended to prevent the repeated entry of a patient. The current hash code that combines the patient's initials, date of birth, and sex has proven to be insufficiently solid, since we found 11 double entries up to now in 1923 records. The entry of the initials has shown to be the most variable and inconsistent factor, since registrations in hospitals often are incomplete or incorrect. A solution may be the entry of a citizen identification number (Dutch: Burger Service Nummer, BSN) using Trusted Third Party (TTP) procedures, which warrants pseudonymization that complies with the strictest privacy legislations. Furthermore, standard procedures for checking data quality will be written, thereby preventing errors in the data entry, such as misspelled gene names and other free text fields. Finally, there currently is no compulsory minimal clinical dataset, which may be very useful for several research questions that rely on detailed clinical data. However, the data requirements will be different for each research question, complicating the definition of the minimal dataset.

Funding

Up to now, the costs for the development of RD5000db have been covered by Dutch blindness foundations. However, once the database is used routinely, it stands to reason that each participating center contributes an annual fee to host, maintain, and update the database. Information about the set-up and maintenance costs for the database can be found in Supplementary Appendix S2.

International Expansion of the Database

Currently, the RD5000db is available for 11 Dutch ophthalmologic centers. In the near future, we want to include data of a

few international centers. The RD5000db thereby has the potential to develop into a pan-European database for the detailed registration of clinical data. By increasing the number of patient records, its potential will grow significantly and facilitate many different studies in the fields of genetics, genotype-phenotype comparisons, and the application of novel therapies.

Acknowledgments

The authors thank Eberhart Zrenner and Eric Troeger (Institute for Ophthalmic Research, Tübingen, Germany) for generously providing the source code of Ophthabase, which was developed as part of the EVI-GENORET Project. Programming services were provided by Raymond H.T. Grootjen, Jeroen van der Grift, Tobias Ph. Poll, Braam Wijsmuller, and Ruud van de Laar of First8 Java Consultancy, Conclusion, Nijmegen, the Netherlands. The authors thank all members of the RD5000 consortium (www.RD5000.nl) for their valuable contributions, in particular, Ymkje Hettinga, Martha Tjon, Meindert de Vries, Anneke I. den Hollander, and Lies H. Hoefsloot.

Supported by the Stichting A.F. Deutman Oogheelkunde Research-fonds, Landelijke Stichting voor Blinden en Slechtzienden, MD Fonds, Stichting Nederlands Oogheelkundig Onderzoek, Rotterdamse Stichting Blindenbelangen, Stichting ter Verbetering van het Lot der Blinden, Algemene Nederlandse Vereniging ter Voorkoming van Blindheid, Oogfonds, Novartis Fonds-Stichting Steunfonds UitZicht. The authors alone are responsible for the content and writing of the paper.

Disclosure: **R.A.C. van Huet**, None; **C.J. Oomen**, None; **A.S. Plomp**, None; **M.M. van Genderen**, None; **B.J. Klevering**, None; **R.O. Schlingemann**, None; **C.C.W. Klaver**, None; **L.I. van den Born**, None; **F.P.M. Cremers**, None

References

- Berson EL. Retinitis pigmentosa. The Friedenwald Lecture. *Invest Ophthalmol Vis Sci*. 1993;34:1659-1676.
- Hartong DT, Berson EL, Dryja TP. Retinitis pigmentosa. *Lancet*. 2006;368:1795-1809.
- Michaelides M, Hardcastle AJ, Hunt DM, Moore AT. Progressive cone and cone-rod dystrophies: phenotypes and underlying molecular genetic basis. *Surv Ophthalmol*. 2006;51:232-258.
- den Hollander AI, Black A, Bennett J, Cremers FP. Lighting a candle in the dark: advances in genetics and gene therapy of recessive retinal dystrophies. *J Clin Invest*. 2010;120:3042-3053.
- Bunker CH, Berson EL, Bromley WC, Hayes RP, Roderick TH. Prevalence of retinitis pigmentosa in Maine. *Am J Ophthalmol*. 1984;97:357-365.
- Hamel C. Retinitis pigmentosa. *Orphanet J Rare Dis*. 2006;1:40.
- Hamel CP. Cone rod dystrophies. *Orphanet J Rare Dis*. 2007;2:7.
- Mockel A, Perdomo Y, Stutzmann F, Letsch J, Marion V, Dollfus H. Retinal dystrophy in Bardet-Biedl syndrome and related syndromic ciliopathies. *Prog Retin Eye Res*. 2011;30:258-274.
- Kalatzis V, Hamel CP, Macdonald IM. First International Choroideremia Research S. Choroideremia: towards a therapy. *Am J Ophthalmol*. 2013;156:433-437.
- Berger W, Kloekener-Gruissem B, Neidhardt J. The molecular basis of human retinal and vitreoretinal diseases. *Prog Retin Eye Res*. 2010;29:335-375.
- Roosing S, Thiadens AA, Hoyng CB, Klaver CC, den Hollander AI, Cremers FP. Causes and consequences of inherited cone disorders. *Prog Retin Eye Res*. 2014;42:1-26.
- Tröger E, Wilke R, Prokofyeva E, Zrenner E. Ophthabase: a generic extensible patient registry system. *Acta Ophthalmol*. 2008;86:0.

13. Talmon JL, Ros MG, Legemate DAPSI: The Dutch Academic Infrastructure for shared biobanks for translational research. *Summit on Translat Bioinforma*. 2008;2008:110-114.

APPENDIX

RD5000 Study Group

Nathalie M. Bax, Carel B. Hoyng, and Stanley Lambertus (Department of Ophthalmology, Radboud University Medical Center, Nijmegen, The Netherlands); Wendy A. van Zelst-Stams (Department of Human Genetics, Radboud University Medical Center, Nijmegen, The Netherlands); Arthur A. B. Bergen (Department of Clinical Genetics, Academic Medical Center, Amsterdam, The Netherlands); José Schuil (Bartiméus, Institute for the Visually Impaired, Zeist, The Netherlands); Mary J. van Schooneveld (Department of Ophthalmology, Academic

Medical Center, Amsterdam, The Netherlands); Laurence Pierrache and Magda A. Meester-Smoor (Department of Ophthalmology, Erasmus Medical Center, Rotterdam, The Netherlands and The Rotterdam Eye Hospital, Rotterdam, The Netherlands); Camiel J. F. Boon (Department of Ophthalmology, Leiden University Medical Center, The Netherlands); Jan Willem R. Pott (Department of Ophthalmology, University Medical Centre Groningen, University of Groningen, The Netherlands); Redmer van Leeuwen (Department of Ophthalmology, University Medical Center Utrecht, The Netherlands); Hester Y. Kroes (Department of Medical Genetics, University Medical Center Utrecht, The Netherlands); Yvonne de Jong-Hesse (Department of Ophthalmology, Free University Medical Center, The Netherlands); F. Nienke Boonstra (Bartiméus, Institute for the Visually Impaired, Zeist, The Netherlands), and Donders Institute for Brain, Cognition and Behaviour, Department Cognitive Neuroscience, Radboud University Medical Centre, Nijmegen, The Netherlands.