THE IIOAB JOURNAL

VOLUME 14 : NO 3 :JULY 2023 : ISSN 0976-3104

Prof. Vasco Azevedo
Federal University of Minas Gerais
Brazil

Editor-in-Chief

Integrative Omics and Applied Biotechnology (IIOAB) Journal Editorial Board:

IIOAB JOURNAL

ARTICLE

DATA MODELING TECHNIQUES FOR DATA WAREHOUSE

Rohith Bharath*

Project Manager, Infosys, Chennai, INDIA

ISSN 0976-3104

ABSTRACT

The Entity-Relationship (ER) model is widely utilized in relational database environments for database design, focusing on day-to-day operations. On the other hand, Multidimensional (MD) data modeling plays a vital role in designing data warehouses aimed at supporting managerial decision-making. It facilitates decision-making by enabling users to delve deeper into detailed information, view summarized information through roll-up operations, select specific items of interest by slicing and dicing dimensions, and reorient the view of MD data through pivoting. When creating a MD model, regardless of whether it follows a star or snowflake schema, the process involves identifying facts, dimensions, and measure attributes. This paper aims to investigate how the Multidimensional model can serve as the primary framework for data warehouse design instead of the ER Model.

KEY WORDS
Entity-Relationship Model; Multidimensional Model; Fact; Dimensions; Attributes

INTRODUCTION

Data modeling is the process of creating a conceptual representation of data, including the relationships between different data entities and the rules that govern them. This representation, called a data model, can be used to design and implement a database [1, 2]. By using data Modeling, organizations can visualize the different types of data used, the connections between pieces of information, and how data is structured and organized. Data Modeling is a method for enhancing data to streamline information flow throughout businesses for varied work purposes [1]. There are several data modeling techniques, in this article we will be focusing on following two models:

- Entity-Relationship (ER) Data Modeling
- Multidimensional Data Modeling

ER MODELING

Entity

A person, place, object, or event of interest to a company or organization is referred to as an entity. A class of objects, or items in the actual world that can be seen and categorized according to their traits and attributes, is represented by an entity [2, 5]. In the [Fig. 1] Customer and Branch are entities.

Relationship

It illustrates the structural relationship and interaction between the model's elements. It can be used to define the connection between two entities [2, 5]. The most occurrences of one entity that can be connected to one instance in another table and vice versa. In the [Fig. 1] Withdraw and Loan are relationships.

Attribute

The traits and properties of an entity are described by its attributes [2, 3, 4]. An entity's attribute names ought to be distinctive and self-explanatory. The minimal cardinality of an attribute is zero when an instance has no value for it, indicating that it is either nullable or optional. If an attribute's maximum cardinality in ER Modeling is more than 1, the modeler will attempt to normalize the entity before elevating the attribute to another entity. As a result, an attribute's maximum cardinality is often 1. In the [Fig. 1] Name, Address and Application are attributes.

DIMENSIONAL DATA MODELING

Dimensional Modeling (DM) is a data structured method designed specifically for data warehouse storage. Dimensional Modeling is used to enhance databases for quicker data retrieval. The "fact" and "dimension" tables that make up the Dimensional Modeling idea were created by Ralph Kimball [2]. A dimensional model is a tool used in data warehouses to read, summarize, and analyze numerical data such as values, balances, counts, weights, etc. Relational models, on the other hand, are designed for the insertion, updating, and deletion of data in an online transaction system that is live [4].

Received: 19 May 2023
Accepted: 04 July 2023
Published: 15 July 2023

***Corresponding Author**
E-mail:
rohithbharath1977@gmail.com

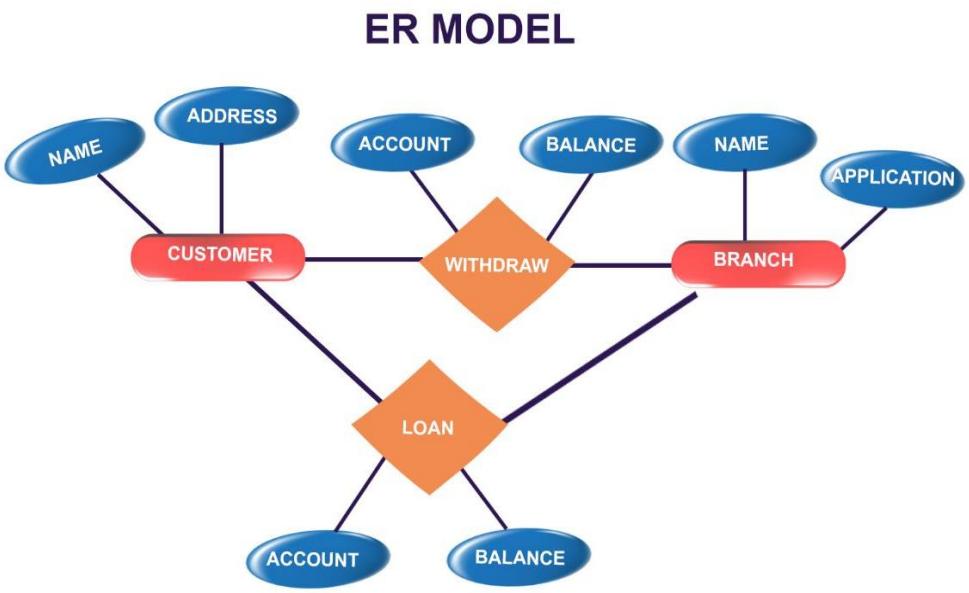


Fig. 1: Entity relationship model.

Dimension

A dimension comprises a set of members or units that share the same type of characteristics [2, 6, 13]. Typically, a dimension is depicted as an axis in a diagram. Within a Multidimensional model, each data point in the fact table is linked to a single member from each of the multiple dimensions. In other words, dimensions provide the contextual framework for the facts. Numerous analytical procedures are employed to measure the influence of dimensions on the facts. Eg. Customer, Location, Sku etc.

Types of Dimensions

In a data warehouse, the dimensions are divided into various categories based on their usage and behavior. Knowing the kind of each dimension before constructing a table in a data warehouse can help you make the best choices. There are various types of dimensions. Major three of them are shown below: conformed dimensions, junk dimensions and slowly changing dimensions.

Conformed Dimensions

Different fact tables can be associated with a conformed dimension while still retaining the same meaning. Conformed dimensions enable cross-domain queries in constellation-type data warehouse systems with several fact tables [7]. The date is an instance of a conformed dimension. Each fact table has the same meaning. The majority of data warehouses contain a single date dimension that is shared by all fact tables as a result. As they must guarantee consistency across several domains, other conformed parameters, which are less evident than the date, may present design difficulties.

Junk Dimensions

In a data warehouse, facts frequently have indicator characteristics like flags, Boolean values, or any other set of values that, due to their low cardinalities, do not make sense as dimensions [8]. A junk dimension is frequently built to combine all of these qualities into one table in order to avoid generating small dimensions for each of these variables and unnecessarily expanding the number and sizes of the fact tables. Instead of keeping the value of each of the attributes in the fact table, it is sufficient to include a single foreign key to the junk dimension table.

Slowly Changing Dimensions

Slowly Changing Dimensions are basically those dimensions whose key value will remain static but description might change over the period of time [9]. For example, the product id in a companies, product line might remain the same, but the description might change from time to time, hence, product dimension is called slowly changing dimension.

Fact

A fact encompasses a set of interconnected data elements, comprising both measures and contextual data. Generally, each fact represents a business entity, a business transaction, or an event that can be utilized in analyzing the business or its processes [10]. Within a data warehouse, facts are incorporated into the central tables that store all the numerical data.
Eg. Sales, Amount, Price

Types of Facts

There are three types of facts:

- ***Additive:*** Additive facts are facts that can be summed up through all of the dimensions in the fact table [2, 11]. Eg. Sales amount.
- ***Semi-Additive:*** Semi-additive facts are facts that can be summed up for some of the dimensions in the fact table, but not the others [2, 11]. Eg. Current balance.
- ***Non-Additive:*** Non-additive facts are facts that cannot be summed up for any of the dimensions present in the fact table [2, 11]. Eg. Profit margin

Hierarchy

A dimension's elements can be arranged into one or more hierarchies. There may be several levels in each hierarchy. Not all members of a dimension locate on the same hierarchy [2, 11]. Eg. Product,Date,Location [Fig. 2].

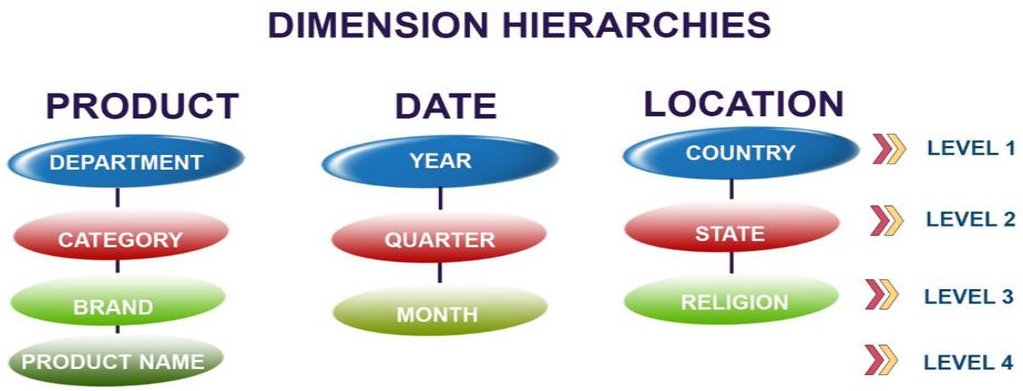


Fig. 2: Product, date and location hierarchies.

Measure

A measure refers to a numerical characteristic of a fact, indicating the performance or behavior of the business in relation to the dimensions [2, 12]. The specific numeric values associated with measures are referred to as variables. A measure is determined based on combinations of dimension members and is situated within the facts themselves. Eg. Sales revenue, Sales volume, Quantity supplied.

TECHNIQUES

Dimensional modeling employs two fundamental models: star schema and snowflake schema.

Star Schema

The Star Schema has the most straightforward structure of all the Schemas. A set of Dimensions Tables are surrounded by the Fact Table in a Star Schema [2, 16]. There are several missing normalizations in these Dimension Tables. The Dimension Tables in this Schema will include a list of characteristics that

characterize the Dimension. Additionally, they have foreign keys that are connected to the Fact Table to produce results [Fig. 3].

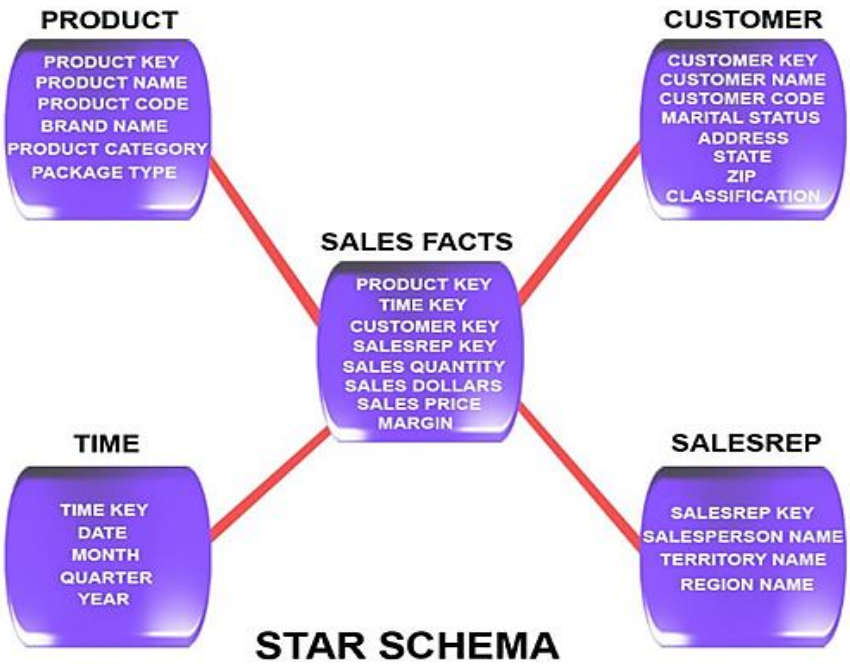


Fig. 3: Star schema.

Snowflake Schema

The Dimensions are fully normalized and divided into additional tables, unlike a Star Schema. Because the data has already been normalized, this Schema takes up less storage space [2, 17]. This Schema's efficient structure makes it simple to add Dimensions, and it also reduces data redundancy [Fig. 4].

Fig. 4: Snowflake schema.

WHY ER IS NOT SUITABLE FOR DATA WAREHOUSES

The ER Model is not easily comprehensible or memorable for end users. It lacks navigation capabilities, and there is no graphical user interface (GUI) that can transform a general ER diagram into a user-friendly format[8, 9]. ER modeling is not well-suited for complex, ad-hoc queries as it primarily focuses on optimizing repetitive and narrow queries. The use of ER modeling techniques undermines one of the key advantages of data warehousing, which is intuitive and high-performance data retrieval. This is because ER modeling often leads to highly normalized relational tables, which can hinder efficient data retrieval.

BENEFITS OF DIMENSIONAL MODEL

Dimensional tables are simpler to interpret than normalized models. The business can easily understand the dimensional model. In order for the business to understand what each fact, dimension, or characteristic implies, this model is built on business terminology. Denormalized and streamlined dimensional models are used for quick data querying. This paradigm is recognized by many relational database platforms, which then optimize query execution plans to improve performance. In a data warehouse, dimensional Modeling produces a high-performance-optimized schema. It results in fewer joins and lessens data redundancy. Models with dimensions can easily adapt to change. More columns can be added to dimension tables without having an impact on currently running business intelligence applications that use these tables.

ER VS MULTIDIMENSIONAL MODEL

The following Table-1 describes the differences between ER and multidimensional modeling.

Table 1: Comparison of ER and dimensional model

ER Modeling	Dimensional Modeling
Entities and Relationships.	Fact Tables and Dimension Tables.
Few levels of granularity.	Multiple levels of granularity.
Real-time information.	Historical information.
It eliminates redundancy.	It plans for redundancy.
OLTP Application	OLAP Application
High transaction volumes using few records at a time.	Low transaction volumes using many records at a time.
Highly Volatile data.	Non-volatile data.
Normalization is suggested.	De-Normalization is suggested.

CONCLUSION

This paper explores the concept of an E-R structured data warehouse without associative entities, specifically fact tables, and discusses its feasibility in light of recent advancements in data warehousing. Several conclusions are derived from the presented arguments. Not all E-R models can be translated into a collection of star schemas that preserve the same information. However, every appropriately designed E-R data warehousing model can indeed be represented as a set of star schemas. Numerous E-R data warehouse models are inadequately constructed because they fail to explicitly acknowledge many-to-many relationships and the necessity to resolve them using associative entities, namely fact tables. Using data warehousing E-R models that only specify atomic data dependency relationships without fact tables can result in poor query response performance in large databases. This, in turn, hinders or even prevents the execution of multi-stage analysis processes. Essentially, it reduces the data warehouse to merely a large staging area for data marts, devoid of its own independent analytical functionality. Considering the emergence of Operational Data Stores (ODSs) and non-queryable centralized staging areas for storing, extracting, cleansing, transforming data, and gathering centralized metadata, the addition of another non-queryable staging area, referred to as a data warehouse, is unnecessary. Instead, what we truly require is a dimensionally modeled data warehouse capable of supporting enterprise-wide Decision Support Systems (DSS). Such a data warehouse should prioritize optimal query response performance and offer advanced OLAP functionality.

CONFLICT OF INTEREST
Authors declare no conflict of interest.

ACKNOWLEDGEMENTS
None.

FINANCIAL DISCLOSURE
None.

REFERENCES

[1] Blakeley JA, N Coburn, PA Larson. [1989] Updating derived relations: Detecting irrelevant and autonomously computable updates. ACM Transactions on Database Systems 14 (3): 369-400.
[2] Kimball R. [1996] The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses. John Wiley & Sons, New York.
[3] Blakeley JA, Coburn N, Larson P. [1986] Updating derived relations: Detecting irrelevant and autonpmously computable updates. In Proceedings of the 12th International Conference on Very Large Data Buses (Kyoto), 457-466.
[4] Firestone JM. [1998] Dimensional Modeling and E-R Modeling In The Data Warehouse White Paper No. Eight. (https://www.dkms.com/papers/dmerdw.pdf, Accessed on March 03, 2023).
[5] Seenivasan D, [2023] ETL (Extract, Transform, Load) Best Practices, International Journal of Computer Trends and Technology, 71(1):40-44,
[6] Hussain H, Malik SUR, Hameed A, et al. [2013] A survey on resource allocation in high performance distributed computing systems. Parallel Computing 39(11):709-736.
[7] Gour V, Sarangdevot SS, Tanwar GS, Sharma A. [2010] Improve performance of extract, transform and load (ETL) in data warehouse. International Journal on Computer Science and Engineering 2(3):786-789.
[8] Korhonen JJ, Melleri I, Hiekkanen K, Helenius M. [2014] Designing data governance structure: an organizational perspective. GSTF Journal on Computing 2(4): 11-17.
[9] Ranjan J. [2009] Business intelligence: Concepts, components, techniques and benefits. Journal of theoretical and applied information technology 9(1):60-70.
[10] Radhakrishna V, SravanKiran V, Ravikiran K. [2012] Automating ETL process with scripting technology. In 2012 Nirma University International Conference on Engineering (NUiCONE), Ahmedabad, India, IEEE, 1-4.
[11] Seenivasan D. [2023] Exploring Popular ETL Testing Techniques, International Journal of Computer Trends and Technology 71(2):32-39.
[12] Ali SMF, Wrembel R. [2017] From conceptual design to performance optimization of ETL workflows: current state of research and open problems. The VLDB Journal 26(6):777-801.
[13] Seenivasan D. [2023] Improving the Performance of the ETL Jobs, International Journal of Computer Trends and Technology 71(3): 27-33.
[14] https://medium.com/ziegert-group/etl-performance-improvement-c5a9bd65b6af (Accessed on March 12, 2023).
[15] https://blog.devart.com/how-to-optimize-sql-query.html Accessed on March 14, 2023).
[16] http://www.ijmer.com/papers/(NCASG)%20-%202013/24.pdf Accessed on March 12, 2023).
[17] https://dataintegrationinfo.com/improve-etl-performance/ Accessed on Jan 03, 2023).
[18] https://medium.com/@data_analytics/etl-for-data-warehousing-1203dc346a4e Accessed on March 14, 2023).

IIOAB JOURNAL

EXPERT OPINION

SECURITY DESIGN CONSIDERATIONS IN ROBOTIC AUTOMATIONS

Sridevi Kakolu*

Technical Architect in Information Technology Department, Boardwalk Pipelines, Houston, TX, USA

ISSN 0976-3104

ABSTRACT

Robotic Process Automation (RPA) is a new favorite among IT leaders. It can be quickly deployed to automate repetitive tasks, and it saves organizations time and money. RPA bots handle sensitive data, moving it across systems from one process to another. If the data is not secured, it can be exposed and can cost organizations millions of dollars. Security design considerations are vital in RPA implementation for its success. Based on my recent Robotic Process Automation project implementation and research, able to put together the design strategies and best practices to implement security in RPA. This paper focuses on RPA introduction, types of RPA, major security risks in RPA process and risk mitigate security design considerations in RPA process implementations for Organizations.

KEY WORDS
Robotic Process Automation, RPA, Unattended Bots, Security in RPA, Best practices for RPA

Received: 03 May 2023
Accepted: 04 July 2023
Published: 12 Aug 2023

INTRODUCTION

Robotic Process Automation is the event-driven software used to automate tasks and processes otherwise performed by humans. In robotic automation, a software robot mimics the work of human users to perform various tasks that are repetitive, high volume, and rules driven. Through automated processes, the software robot executes a workflow involving multiple steps and interactions with different enterprise applications. Robotic process automation remains a popular software market for improving operational efficiency with tactical automation.

TYPES OF RPA

Robotic Process Automations can be executed in two modes:

- Attended
- Unattended

An attended automation assists an agent/human in handling simple, repetitive tasks. In contrast, an unattended automation automates specific tasks which do not require agent/human intervention [Table 1]. Depending on the project or use case, the robotic automations can interact with enterprise applications, databases, or financial systems. RPA can help process the required tasks with or without presence of a human.

As per 2022 Gartner Magic Quadrant Evaluation Report, strategic planning assumptions: "By 2024, 95% of RPA vendors will offer automation via both API and UI integration. By 2024, 80% of enterprise customers who have deployed attended automation primarily on a desktop will pivot to wider UX covering web, mobile and voice interfaces." [1]

Table 1: Types of Robotic Automations.

Attended Automation	Unattended Automation
Executes workflow on an end user's desktop to assist in human work	Runs on a virtual or server and does not require human intervention
Works alongside agents to improve productivity and quality	Runs on a dedicated workstation to execute fully automated processes
Automates 20-90% of a given task	Automates 100% of given tasks

BENEFITS OF RPA

Over the past few years, robotic process automation (RPA) has become a popular technology. This is due to its ability to automate repetitive and high-volume tasks in order to reduce manual effort, eliminate error

***Corresponding Author**
E-mail:
Sridevi.Kakolu@gmail.com

and improve process productivity. With RPA, software bots can mimic human actions such as logging into various applications/systems. They can also navigate through user interfaces to perform tasks such as creating tickets and downloading data. Bots can provision and deprovision user access and respond to customer queries. RPA is versatile and flexible, allowing it to integrate easily with existing processes. It helps reduce cost, maintain consistent quality, improve delivery timelines, and enhance the customer experience [Fig. 1].

Fig.1: Benefits of RPA.

SECURITY RISKS WITH RPA

Organizations looking to implement RPA should be aware of the security-related risks [Fig. 2]. These include:

Compromise privileged access accounts

In terms of RPA security, the risks of privileged access abuse by RPA bots are similar to those by humans. For example, privileged access given to an RPA bot account may be used by attackers. They may break into the system and steal or misuse sensitive business information. For better auditing and troubleshooting, it is essential to distinguish bot activities from those of employees. Never use an employee's credentials for RPA implementation. Create unique identities for every bot in the system, and do not store passwords in the source code. Keep passwords in a centralized, encrypted location such as a password vault and change them frequently. Limit the number of employees who have access to RPA credentials. Configure a robust authentication method like two-factor authentication or token authentication for extra security.

Malfunctioning and system outage risks

System outage (or downtime) refers to the period when a system/network cannot perform its primary function. Downtimes can happen because of numerous reasons. The most frequent reasons for this issue are human error, outdated or unstable hardware/software, bugs in the server operating system, and integration/interoperability issues. In RPA, there are two potential risk scenarios related to system outage. First, unexpected network failure may disrupt the bot's operation leading to a significant loss in productivity. Second, a rapid sequence of bot activities may cause system failure or outage. For instance, in 2018 on Amazon Prime day, millions of shoppers faced a high-profile outage on the Amazon "Deals" page because its servers did not manage such a massive online traffic spike.[2]

Data breach

Confidential information is any information related to a company's business and affairs that is not available to the public and has commercial value. Unauthorized disclosure of financial information, marketing plans, upcoming projects, and other confidential materials may have devastating consequences. In RPA, disclosure of confidential information may occur with the intentional or negligent improper training of an RPA bot. This causes leakage of confidential data, such as payment or credit card data, to the web.

System vulnerabilities

Vulnerabilities are weaknesses in an information system that allows cyber attackers to illegally gain access to the system. One of the ways vulnerabilities may appear is when a malicious user behaves imprudently by visiting an unsafe website. In this case, the website is a threat resource that triggers vulnerability. Some of the most common examples of vulnerabilities are: missing data encryption, SQL injection, missing authorization, cross-site scripting and forgery, weak passwords, upload of infected software. Even though advanced RPA systems nowadays use encryption while transferring data, there are still low-security-level RPA tools. Here, non-encrypted data transfer may cause sensitive data leakage.

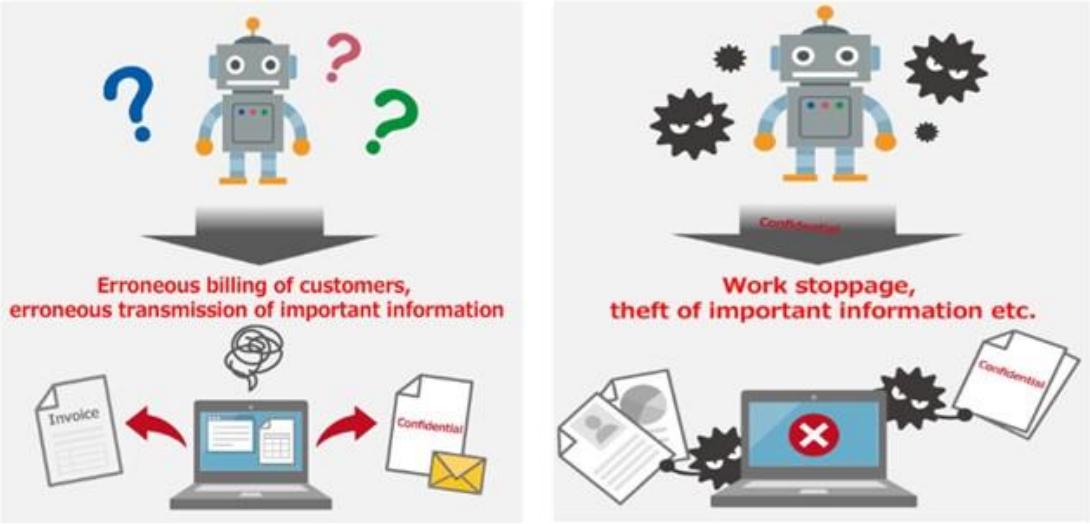


Fig. 2: Security risks in RPA.

Lack of Visibility on the Bots executions

Audit logs capture bot activity, and these logs are important to track bot health and effectiveness. For instance, if a bot stops working, the audit log helps identify the underlying reason. The reason may be improper use by an employee or malicious code. Bots need to be periodically monitored at various levels to ensure they do not misbehave. Misbehavior can lead to high error rates and potential damage. In some cases, bots may not perform as intended due to erroneous coding or inadequate testing. This will result in issues and errors during go-live. Besides RPA software out of the box logging, automation logic design needs to take care of detailed logging while executing set of actions. Audit logs can be monitored in dashboards also configure notifications, then there will not be much visibility on any of these issues.

SECURITY DESIGN CONSIDERATIONS IN RPA

Robotic automations execute the logic based on its design and implementation. This means it has the potential to touch every enterprise application within the organization and the confidential personal and customer data within it. Whether it is an attended automation or unattended automation, it is important to consider security standards within the Robot automation design. The cost of a security incident can be tremendous. Enterprise-ready RPA must assure both business and IT that the RPA deployment will not compromise security or compliance [Fig. 3].

Below five best practices are helpful for reliable and secured RPA design:

- Accountability for Bot Actions
- Automating Credential Management
- Strong Governance Framework
- Continuous Review and Change Control
- Logging, Auditing and Monitoring

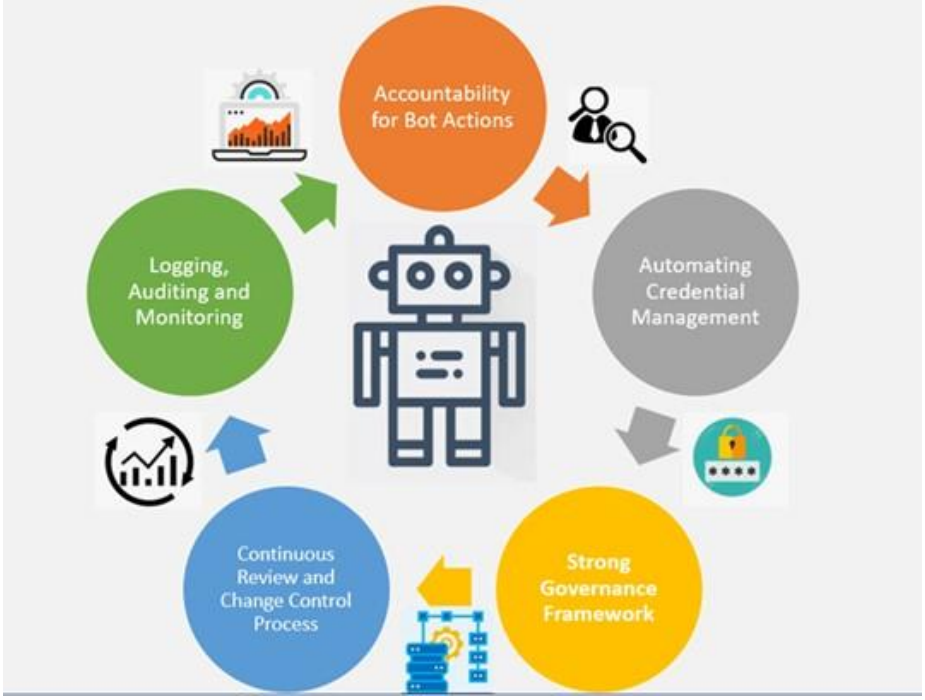


Fig. 3: Security design considerations in RPA.

Accountability for Bot Actions

For better auditing and troubleshooting purposes, it is essential to distinguish the activities of a bot from those of an employee. Never use an employee's credentials for RPA implementation. Create unique identities for every bot in the system, and do not store passwords in the source code. Keep passwords in a centralized, encrypted location such as a password vault and change them frequently. Limit the number of employees who have access to RPA credentials. Configure a robust authentication method like two-factor authentication or token authentication for extra security.

Automating Credential Management

Successful RPA deployments require automated credential management, including machine-generated passwords, automatic password rotation, identity verifications and just-in-time or time-limited credential access. RPA teams can save passwords in single password storage or vault without creating any security leaks. Never use an employee's credentials for RPA implementation. IT administrators can configure minimum access rights for a bot to access applications and databases.

Strong Governance Framework

It is very important to define rules and regulations to maintain security in RPA solutions. Without proper governance, RPA cannot ensure the security it is supposed to offer. Detailed criteria, development criteria, and business justification are some features that fall under an excellent governance framework.

- **Roles & Responsibility Management** - Build and implement a system with clear roles and responsibilities. Roles should include everyone in the department/team responsible for the automation process.
- **Strategy and Regulations** - The company should clearly elaborate the rules and requirements set out in their current security regulations. They should also provide adequate supervision to ensure compliance.

- **Awareness** - Top managers should raise awareness of RPA-related risks and the potential impacts. Awareness should be spread internally (within the responsible teams) and externally (among the RPA bots' creators). Regularly validate RPA scripts and audit logs to ensure a bot is working correctly.

Vendor and internal teams should work together to establish a robust governance framework. The framework must clearly define the automation scope, prioritize identified RPA candidates, and evaluate regulatory and business risks for each RPA candidate. The framework needs to define each team member's roles and responsibilities clearly. It is also advisable to update the company's Information Security Management System (ISMS) and Identity and Access Management (IAM) policies to incorporate RPA specific requirements.

Continuous review and change control

Create a transparent business continuity plan that specifies the backup procedures and data sources required to carry out every task. It is the responsibility of an internal audit team to check and review the documents in the business continuity plan to see if there is any information, like how to restart each process/activity even after failure. Build weekly or monthly review plans on overall RPA infrastructure in the company and review Bot performance. Implement framework-based design approaches to maintain the Bot's logic easily by the developers. Implement CI/CD pipeline process to deliver RPA software upgrades or patch updates and deliver fixes smoothly.

Logging, auditing and monitoring

Enforce proper regulations to monitor the performance of RPA bots and ensure that all bots function in accordance with the set rules. Periodic risk assessment is necessary to track the possibilities of new risks, mitigate, and review security risks in the RPA, to check if any restrictions have been lifted, and to determine if any RPA bot needs to be avoided. It is critical to monitor and log every transaction of an RPA script. Efficient security and risk management practices ensure consistent and accurate logging. It is a good practice to secure RPA logs in a separate system and encrypt sensitive data. Rapidly detect and respond to unauthorized or anomalous robot behavior by assigning human managers, enforcing least privilege, and making actions traceable.

RPA SECURITY DESIGN CHECKLIST

Above 5 best practices action plan helps Security and Risk Management leaders to mitigate RPA risks. Below are a few more comprehensive security checklists that can be useful when starting to design and implement RPA. [3]

- Enhance software development practices to include secure bot development/deployments.
- Implement bots based on input feed from a secured location.
- Treat a robot like a user and create a separate set of credentials.
- Implement integrations with a secret server using access token mechanism to get credentials.
- Maintain a password vault to store bot credentials and rotate bot credentials often.
- Establish mechanisms to find, avoid, and control bot abuse such as a provision to lock down bots.
- Do not leave any credentials in the source code.
- Use two-factor authentication for an extra layer of security.
- Follow the principle of least privilege and grant only the necessary permissions to the bots.
- Ensure that all transactions are correctly logged.
- Implement email notifications in the bot logic for failures and successful processing.
- Review RPA scripts and logs regularly.

CONCLUSION

Organizations adopting RPA to improve productivity should plan their implementations carefully to protect themselves from security breaches. RPA creates new application layers that are vulnerable to risk. Moreover, without constant supervision, bots may not work effectively, causing issues, errors, and potential damage. Since bots may need access to confidential information, it is imperative for organizations to institute the right security measures. Some of these measures include creating governance frameworks, audit logs, password vaults, and version controls. Establishing these processes will allow RPA to manage security risks by itself. This ensures best bot performance and reduced business risk.

CONFLICT OF INTEREST
Author declare no conflict of interest.

ACKNOWLEDGEMENTS
None.

FINANCIAL DISCLOSURE
None.

REFERENCES

[1] https://www.gartner.com/smarterwithgartner/4-steps-to-ensure-robotic-process-automation-security (Accessed on 5 April, 2023)
[2] https://www.uipath.com/resources/automation-analyst-reports/10-golden-rules-of-rpa-success-forrester-report (Accessed on 4 April, 2023)
[3] https://academy.pega.com/topic/pega-robotic-automation-introduction/v2/in/2976/5406 (Accessed on 3 April, 2023)
[4] https://www.xenonstack.com/blog/rpa-security-checklist (Accessed on 3 April, 2023)
[5] https://www.infosys.com/services/cyber-security/documents/rpa-security.pdf (Accessed on 4 April, 2023)
[6] https://electroneek.com/blog/security-concerns-in-rpa-4-step-guide-to-address-them/ (Accessed on 5 April, 2023)

IIOAB JOURNAL
ISSN: 0976-3104

ARTICLE

GENE THERAPY OF PDE6B-RETINOPATHY AS THE PHOTOTRANSDUCTION CYCLE MOLECULE: RELEVANCE AND THE MOST PROMISING TREATMENT METHODS

Varvara Safonova[1*], Vera Kopylova[2]

[1]*Moscow State University, Bioengineering and Bioinformatics Faculty, Moscow, RUSSIA*

[2]*Summer school "Letnyaya shkola", Moscow, RUSSIA*

ABSTRACT

Retinitis pigmentosa (RP) is a group of inherited retinal disorders characterized by the progressive photoreceptors and pigment epithelial cells dysfunction. It is the most common genetic retinal degeneration, responsible for loss of vision of most young people worldwide. Symptoms of RP include deterioration of vision in the dark, a decrease in peripheral vision ending up in tunnel vision. More than 50 different genes are involved in the development of RP. The products of these genes are involved in various processes of the visual and phototransduction cycles or are structural elements of the retina. In this paper we focused on one clinical example PDE6B-associated retinitis pigmentosa and analysis of the vision loss pathogenesis, gene network regulation and possible treatment approaches.

KEY WORDS
Retinitis pigmentosa, PDE6B, transcriptome analysis phosphodiesterase, small-molecule, treatment

Received: 3 June 2023
Accepted: 20 June 2023
Published: 10 July 2023

***Corresponding Author**
Email:
varyaSafonov@yandex.ru
Tel.: +7 920 001 53 32

INTRODUCTION

Retinitis pigmentosa (RP) is a group of inherited retinal disorders characterized by the progressive photoreceptors and pigment epithelial cells dysfunction. It is the most common genetic retinal degeneration, responsible for loss of vision of most young people worldwide [1]. Symptoms of RP include deterioration of vision in the dark, a decrease in peripheral vision ending up in tunnel vision. More than 50 different genes are involved in the development of RP [2]. The products of these genes are involved in various processes of the visual and phototransduction cycles or are structural elements of the retina.

Visual processes are based on a phototransduction mechanism that converts the primary light signal in photoreceptor cells. The quantum of light is absorbed by 11-cis-retinal and isomerizes it into a fully trans-form. This is the only reaction that depends on light. The cis-trans retinal transition causes a conformational rearrangement of the protein part of the rhodopsin (opsin) molecule. As a result, rhodopsin acquires the ability to interact with the next protein in the phototransduction cycle. Transducin belongs to the family of heterotrimeric G-proteins and consists of three subunits. Active transducin activates the next protein of the visual cascade-heterotetrameric phosphodiesterase (PDE) of cyclic GMF (cGMP). This enzyme hydrolyzes cyclic guanosine monophosphate (cGMP) at a high rate. A drop in the concentration of free cGMP in the cytoplasm leads to hyperpolarization of the cell membrane. This electrical potential is a photoreceptor signal, which is then transmitted in the first synapse of the retina to the next nerve cells. Violation of any of the links in this chain can lead to incorrect operation of photoreceptor cells and, as a result, vision disorders.

MATERIALS AND METHODS

Oftalmic Russian IRD (inherited retinal disorders) clinical database was searched for the visual cycle and phototransduction gene mutations cases and out of 118 cases a male 18 y.o patient case with hemizygous mutation in PDE6B gene c.1580T>C p.Leu527Pro NM_000283.3 (chr4:654368T>C) was selected for further analysis. We used mouse animal model for transcriptome analysis. Total RNA was extracted from Pde6b-/- retina in our experiment using RNA Qiagen kit and the RNA quality and quantity was assessed by the absorbance at 260 nm/280 nm using a NanoDrop ultraviolet spectrophotometer; RNA integrity was verified by 1% agarose gel electrophoresis. The cDNA library was purified, and the library was sequenced on a NovaSeq 6000 platform. RNA-sequencing reads were first trimmed to remove poly(A) and unqualified reads with Cutadapt (v1.15), then aligned to the Homo sapience GRCm38 genome using Ensembl. The counts were summarized at the gene level using HTseq (0.9.1). Gene expression values were calculated from fragments per kilo basis per million fragments (FPKM). These FPKM values were used to generate a table with the Prism software. Paired differential gene expression analyses were performed with DEseq (1.30.0) with screened conditions as follows: multiple expression difference|log2fold change| > 1 and p-value < 0.05. The main biological functions associated with the differentially expressed genes were determined by GO (p-value < 0.05). The KEGG pathway enrichment

IIOAB JOURNAL
ISSN: 0976-3104

analysis of the differential genes focused on the associated enriched pathways (p-value < 0.05). A protein–protein interaction (PPI) analysis of differential genes was performed using the STRING database (http://string-db.org) and genemania.org webtool to reveal the relationships between the target genes. The PPI network model was generated using Cytoscape. Statistics 28 software was used to analyze results.

RESULTS

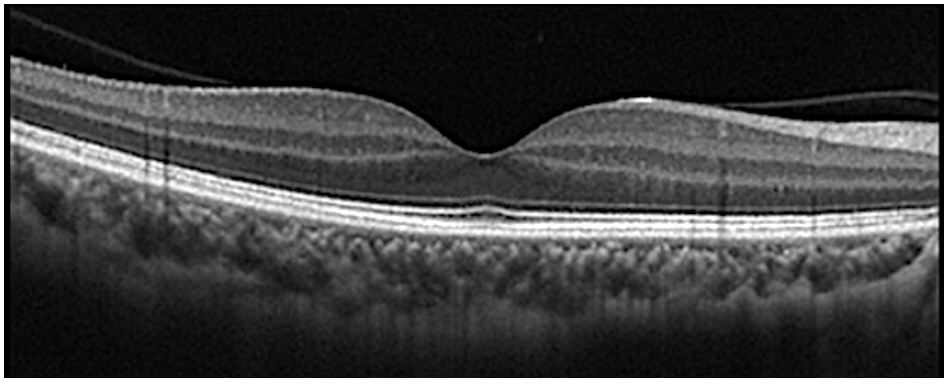
This paper discusses the PDE6B gene encoding the β-subunit of cGMP phosphodiesterase. This protein is involved in the transmission and amplification of visual signals and is necessary for the formation of a functional phosphodiesterase holoenzyme. Phosphodiesterase itself is an effector protein in the phototransduction cycle and catalyzes the conversion of cGMP to GMF, which causes the closure of ion channels and a decrease in the level of glutamate in rods.

To date, there is no single treatment for RP, but thanks to recent advances in imaging technology, DNA sequencing, gene therapy and stem cell biology, the number of clinical trials has increased,

Most mutations in the phosphodiesterase gene result in improper regulation of cGMP levels and disruption of ion channels. It is believed that this eventually causes the death of receptor cells. Gene therapy effectively restores long-term retinal function and vision and offers great prospects for human treatment. The most promising methods in the treatment of retinitis pigmentosa today are CRISPR/Cas9 and AAV gene therapy.

A 18 y.o. male patient with the hemizygous mutation in PDE6B gene c.1580T>C p.Leu527Pro NM_000283.3 (chr4:654368T>C) presented symptoms of night blindness at the age of 4 y.o, his BCVA (best corrected visual acuity) was 0.3/0.4 OD/OS at the first examination. subsequently his visual fields narrowed over the course of 5 years, his BCVA was relatively stable. His OCT (optical coherence tomography) scans are presented on [Fig. 1B].

A



B

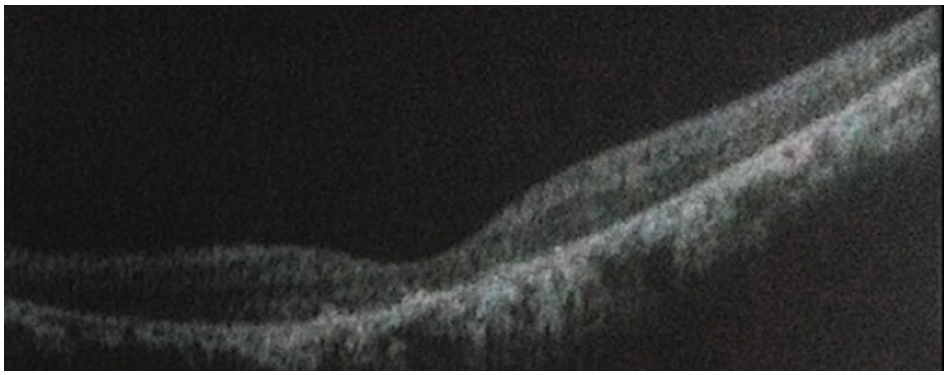


Fig. 1: OCT (optical coherence tomography) scans of a right eye healthy control retina (A) and a patient with *PDE6B* selected mutation (B).

PDB (international protein database) of P35913 UniProt identifier for PDE6B was analyzed for the proven pathogenic mutation c.1580T>C influence on protein structure in [Fig. 2].

IIOAB JOURNAL
ISSN: 0976-3104

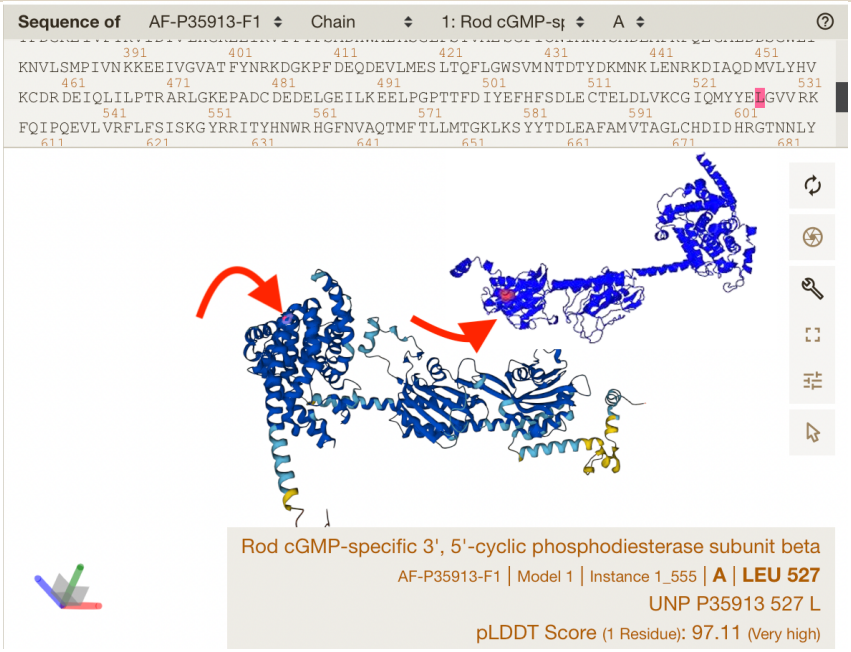


Fig. 2: The structure of phosphodiesterase 6 beta protein. The site of the mutation is marked in red PDE6B c.1580T>C p.Leu527Pro NM_000283.3 (chr4:654368T>C).

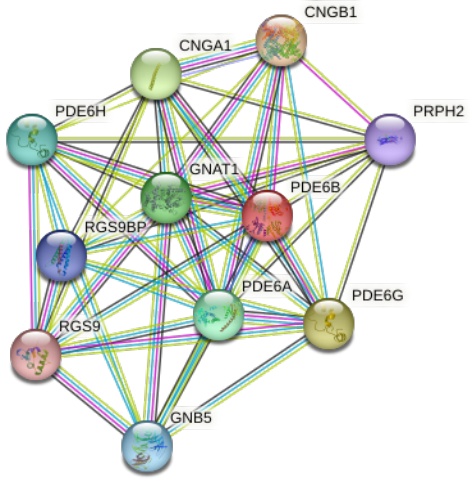


Fig. 3: *PDE6B* gene and protein interaction network involves mostly molecules of visual cycle and phototransduction.

Table 1: Transcriptome analysis of the PDE6B gene and protein network. We analyzed RNA expression in the mutated animal and compared it to normalized wild-type gene expression pattern. Expected level of expression calculated by in silico algorithms is presented in expected expression column and the experimental data analysis is presented in the experimental expression column.

Gene name	Gene function explanation	Expected expression	Experimental expression
Visual cycle			
RHO	Rhodopsin in rods	↔	↔
OPN1LW/MW/SW	Opsin 1 (cone pigments), long/medium/short-wave-sensitive	↓	↔
RBP1/2/3/4	Retinol binding protein subunits, cellular	↓	↓
RLBP1	Retinaldehyde binding protein 1	↔	↓
RPE65	Retinal pigment epithelium-specific protein 65kDa	↓	↑
LRAT	Lecithin retinol acyltransferase	↔	↔
RDH5/8/10/11/12/16	Retinol dehydrogenase different types	↓	↔
ABCA4	ATP-binding cassette, sub-family A (ABC1), member 4	↓	↑
Phototransduction			
GNAT1	Transducin, rod-specific	↓	↑
GNAT2	Transducin, cone-specific	↓	↑

IIOAB JOURNAL
ISSN: 0976-3104

PDE6A	Phosphodiesterase 6 alpha	↓↓	↑↑
PDE6B	Phosphodiesterase 6 beta	↓↓	↓↓
PDE6G	Phosphodiesterase 6 gamma	↓↓	↑↑
CNGA1, CNGA3, CNGB1, CNGB3	Cyclic nucleotide-gated channel subunits, cone and rod specific	↔	↔
GUCA1A/B/C	Guanylate cyclase activator 1B (retina)	↑↑	↑
GUCY2D/2F	Guanylate cyclase 2D/2F, membrane (retina-specific)	↑↑	↓
Secondary			
NRL	Neural retina leucine zipper	↓	↓
NR2E3	Nuclear receptor subfamily 2 group E	↓	↓
CRX	Cone-rod homeobox	↔	↔

We analyzed PPI (protein-protein interaction) in String tool [Fig. 3] and created a table 1 of the genes for transcriptome analysis followed by the graphic [Fig.4] with the comparison of the expression levels in normal (blue) and mutated (orange) cases.

Analysis of expression shows unexpected rise of RPE65, ABCA4, GUCA1B and GUCA1C and down-regulation of GUCY2D, and GUCY2F proteins. This data needs further verification, but initially we can tell that both Phosphodiesterase 6 alpha/gamma and Guanylate cyclase activators 1B/1C were up regulated in case of mutation. RPE65 and ABCA4 levels were unexpectedly high, also both rods and cones transducin levels were higher than expected. We expected Guanylate cyclase expression to increase, and it was decreased which shows some additional mechanisms of regulation in the live cells comparing to isolated molecular analysis.

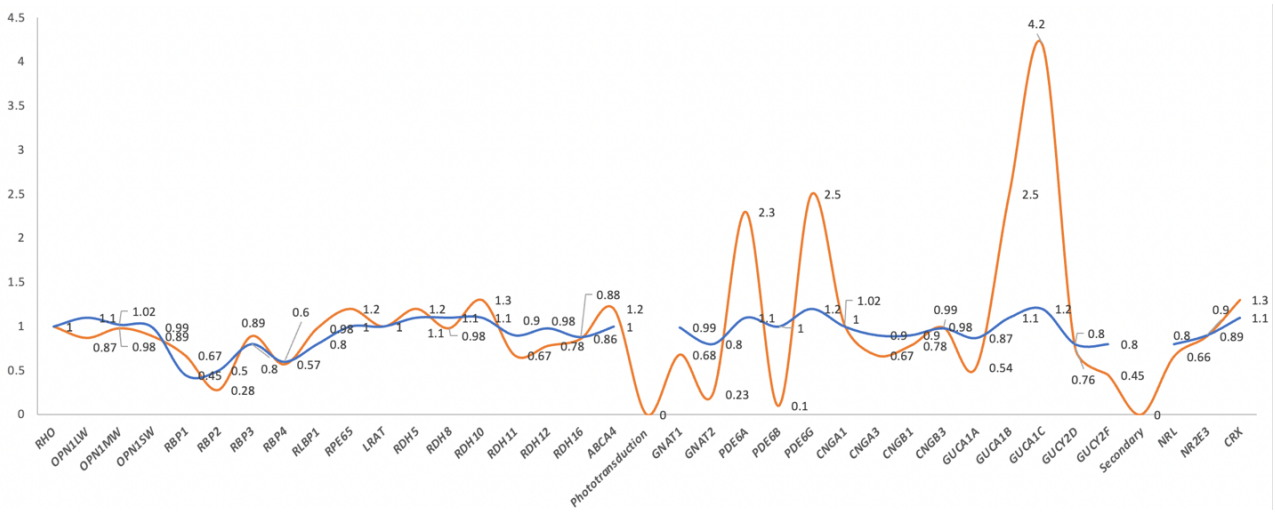


Fig. 4: Transcriptome analysis of the PDE6B-related proteins.

DISCUSSION

In mice, a nonsense mutation was found in the pde6b gene, which leads to incorrect operation of one of the phosphodiesterase subunits. CRISPR/Cas technology was successfully applied to mice with this mutation. This predicts excellent results for gene editing in diseased human tissues, since Pde6b, a mutated gene in mice, has orthologous intron-exon relationships comparable to the human PDE6B gene [3].

Viral vectors are a modern tool for delivering genetic material to the cell. Their main advantages: the ability to integrate the target gene into the desired location of the host genome, which prevents unwanted mutations; embedding in both dividing and resting cells; a wide transduction profile; low immune response; strong and stable expression of the transgene. The essence of the method is the penetration of the virus into the target cell, where the genome is then expressed with the necessary gene inserted into it (in our case, the phosphodiesterase gene).

Experiments have shown that the introduction of adeno-associated viral vectors can prevent retinal degeneration in mice, which is reflected in significant structural, biochemical, and electrophysiological changes. It is worth noting that clinical trials have already reached the second stage after being successfully tested on dogs. These results serve as the basis for studying the long-term rescue of the retina in humans in the future [4][5].

ISSN: 0976-3104

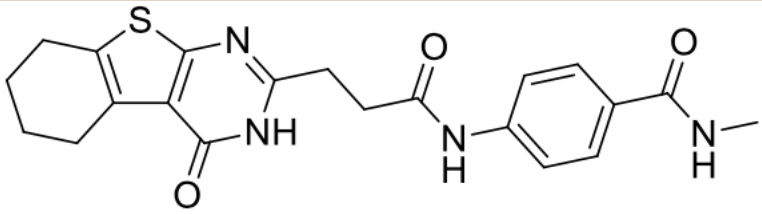


Fig. 5: *Analogue of acrylamide-azobenzene-quarternary ammonium (AAQ) molecule that can be activated by 380 nm blue light (cis-form) and deactivated by 500 nm green light (trans-form).*

Also, it is worth mentioning small molecules treatment approaches that can potentially up-regulate visual cycle in the presence of abnormal PDE6B product. In case of PDE6B mutation the signal is not amplified. What are the possible ways of signal amplification? Some researchers suggested using photo switch molecules [6], for instance, acrylamide-azobenzene-quarternary ammonium (AAQ) molecules [Fig. 5] to serve as a photo activated/deactivated molecule to fire ganglion cells in the case of lack of signal activation. Such kind of small molecule can serve as a molecular channel blocker in light activated *trans*-form and can release the channel in the *cis*-form.

CONCLUSION

One clinical example PDE6B-associated retinitis pigmentosa with robust molecular investigation and transcriptome analysis of the genes and proteins involved in photocycle and phototransduction, gene network regulation and possible treatment approaches of gene therapy, gene editing and small molecules treatment were discussed in this paper.

ETHICS STATEMENT

This research adhered to the tenets of the Declaration of Helsinki. Patient signed informed consent form for the scientific analysis.

CONFLICT OF INTEREST

Authors declare no conflict of interest.

ACKNOWLEDGEMENTS

Authors would like to thank Irina Zhegulina, MD, PhD for mentorship and valuable comments during project preparation and Marianna Weener, MD, PhD for scientific guidance and providing the access to the IRDs registry database.

FINANCIAL DISCLOSURE

None

REFERENCES

[1] Pagon RA. [1988] Retinitis pigmentosa. Surv Ophthalmol.33: 137–177.
[2] Retinitis Pigmentosa | National Eye Institute. [cited 24 Jul 2020]. Available: https://nei.nih.gov/health/pigmentosa/pigmentosa_facts
[3] Wu WH, Tsai YT, Justus S, Lee TT, Zhang L, Lin CS, et al. [2016]CRISPR Repair Reveals Causative Mutation in a Preclinical Model of Retinitis Pigmentosa. Mol Ther.;24. doi:10.1038/mt.2016.107
[4] Pang J-J, Boye SL, Kumar A, Dinculescu A, Deng W, Li J, et al.[2008] AAV-Mediated Gene Therapy for Retinal Degeneration in the rd10 Mouse Containing a Recessive PDEβ Mutation. Invest Ophthalmol Vis Sci.49: 4278–4283.
[5] Prichard V, Provost N. [2016] AAV-mediated Gene Therapy Halts Retinal Degeneration in PDE6β-deficient Dogs.
[6] Drivas TG, Bennett J. [2012]The bionic retina: a small molecule with big potential for visual restoration. Neuron. Jul 26;75(2):185-187. doi: 10.1016/j.neuron.2012.06.010.