



# Rare Disease Data Center (RDDC)



AI Meets Bioinformatics:  
Revolutionizing Rare Disease Research



Your One-Stop Solution for  
Gene, Disease, and Model Data



Guangzhou Rare Disease  
Gene Therapy Alliance



清華珠三角研究院  
Research Institute of Tsinghua, Pearl River Delta



cyagen

# What is the Rare Disease Data Center?



The Rare Disease Data Center (RDDC) is a cutting-edge database co-developed by the AI Innovation Center of Tsinghua Pearl River Delta Research Institute and Cyagen Biosciences. It integrates open-source data from domestic and international sources, including epidemiology, drug development, disease-related gene maps, mutation sites, and rodent models. By combining genetic data resources with AI and bioinformatics technologies, RDDC deploys tools such as the Pathogenicity Predictor and RNA Splicer, significantly advancing rare disease research.

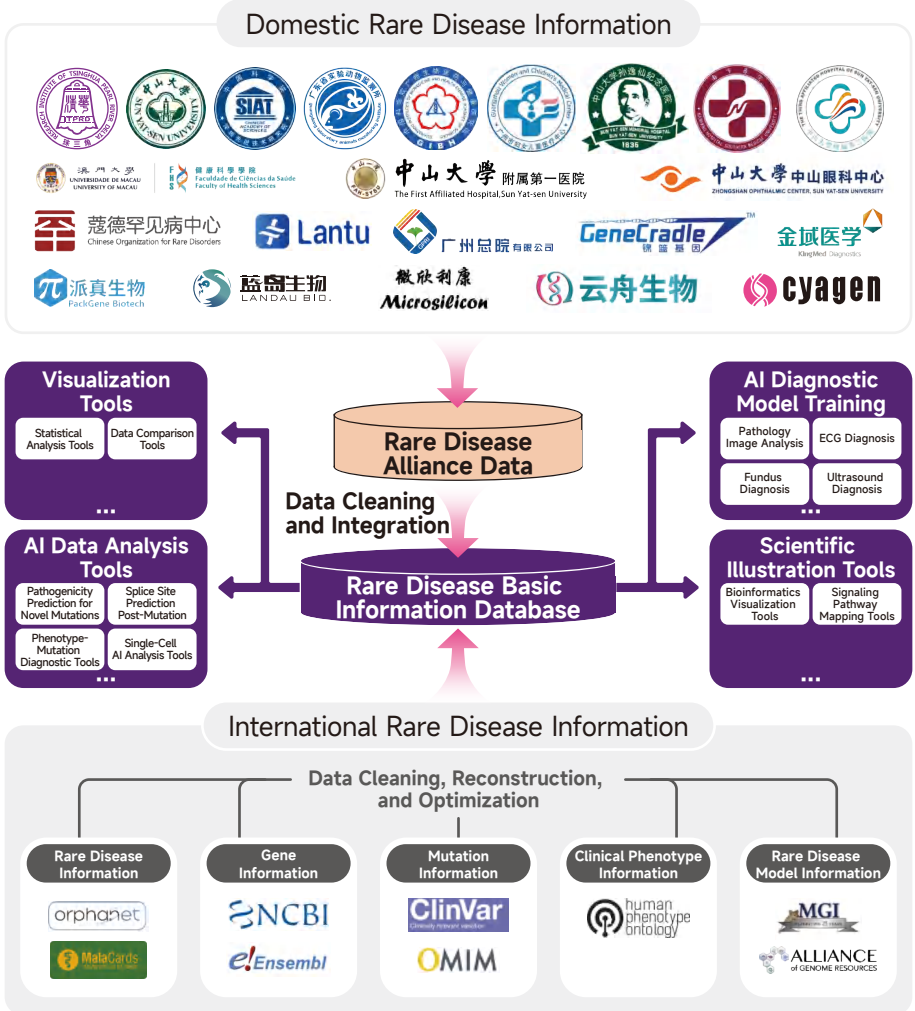


Figure 1. Data Sources and Applications



## Comprehensive Information Retrieval



### Genes

RDDC contains information on 64,099 genes across humans, mice, and rats.



### Diseases

RDDC integrates data from Malacards, OMIM, Orphanet, ClinVar, and the Rare Disease Alliance, covering 21,933 diseases.



### Mouse Models

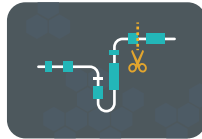
RDDC includes 107,541 gene-edited mouse models from various literature sources.

## AI Prediction Tools: From genomics to proteomics, AI empowers life science research, helping you explore uncharted territories.

### Pathogenicity Predictor



### RNA Splicer



### AI Assistant



## Bioinformatics Analysis Tools: Fast, accurate, and intelligent tools to decode data and accelerate life science research.

### ASO Predict



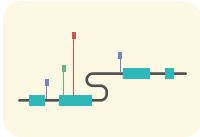
### DNA Reverse Complement Tool



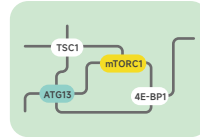
### Sequence Alignment



### Variant Viewer



### Pathway Analysis



# Core Function Use Cases



## Diseases & Mutations

Genes are the building blocks of heredity, carrying instructions that define traits and functions. Mutations in these genes can disrupt protein function, leading to disease. Understanding how genetic mutations relate to diseases is essential for developing targeted therapies that can improve patient outcomes or even cure diseases.

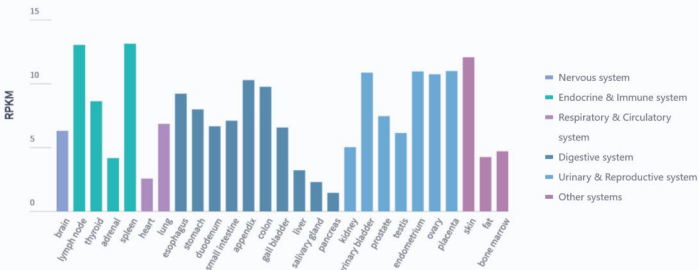
#	Disease	Anatomical Category	Score <sup>+</sup>	Mutations <sup>+</sup>
1	<a href="#">Li-Fraumeni Syndrome (LFS)</a>		1764.42	<a href="#">2052(28996)</a>
2	<a href="#">Papilloma of Choroid Plexus (CPP)</a>		1464.26	<a href="#">30(428)</a>
3	<a href="#">Osteogenic Sarcoma (OSRC)</a>		1376.1	<a href="#">30(434)</a>
4	<a href="#">Glioma Susceptibility 1 (GLM1)</a>		1109.43	<a href="#">33(467)</a>
5	<a href="#">Basal Cell Carcinoma 7 (BCC7)</a>		1107.87	<a href="#">30(428)</a>
6	<a href="#">Adrenocortical Carcinoma, Hereditary (ADCC)</a>		1100.02	<a href="#">30(428)</a>
7	<a href="#">Bone Marrow Failure Syndrome 5 (BMF5)</a>		1057.85	<a href="#">26(354)</a>

## Gene Expression

Gene expression measures how active genes are in specific tissues or cells. These patterns are crucial for understanding biological functions and the mechanisms behind diseases. Studying gene expression helps uncover links between genes and diseases, offering new avenues for prevention and treatment.

### Tissue-specific RNA expression

Organ Abundance Alphabetical



### Cell-specific RNA expression

Organ Abundance Alphabetical

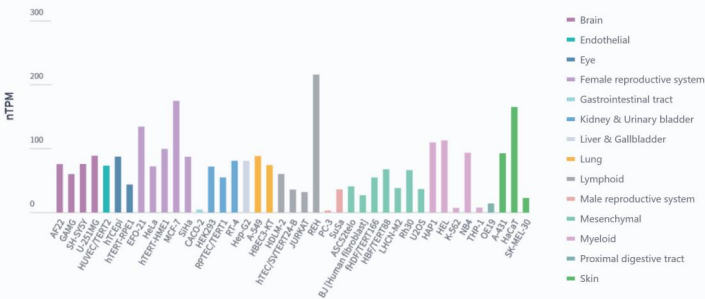


Figure 2. Rich structured data and visualizations.



RESEARCH

Open Access

De novo variations of ANK1 gene caused hereditary spherocytosis in two Chinese children by affecting pre-mRNA splicing



RESEARCH

Open Access

New Chinese databases are a boost for rare-disease science



Figure 5. A powerful AI-assisted diagnostic tool, cited in multiple research papers.

# AI-Assisted Prediction Case Study: Duchenne Muscular Dystrophy (DMD)



## Clinical Features

Duchenne Muscular Dystrophy (DMD) is an X-linked recessive disorder caused by mutations in the DMD gene, the largest gene in the human genome (2.5 Mb). Approximately 60-65% of cases result from large deletions of one or more exons, while 20% are caused by single nucleotide variants, including frameshift, nonsense, missense, and insertion-deletion mutations. In-frame mutations typically lead to Becker Muscular Dystrophy (BMD), a less severe form, whereas frameshift mutations result in severe DMD phenotypes. Whole-genome sequencing of patient samples revealed a hemizygous mutation in the DMD gene: c.4675-2A>G, causing aberrant splicing.

Category	Details	Category	Details
Gene	DMD	Normal Population Frequency	--
Chromosomal Location	chrX: 32398799	ACMG Pathogenicity Analysis	Pathogenic
Transcript/Exon	NM_004006; Exon34	Disease/Phenotype	Duchenne Muscular Dystrophy (XLR) Becker Muscular Dystrophy (XLR) Dilated Cardiomyopathy 3B (XL)
Nucleotide/Amino Acid	c.4675-2A>G	Variant Origin	Spontaneous
Homozygous/Heterozygous	Heterozygous		

Table 1. Whole Exome Sequencing Results for DMD Patients

## ■ AI Prediction Results

Using RDDC's AI tools—Pathogenicity Predictor and RNA Splicer—the pathogenicity and RNA splicing effects of the mutation were predicted.

### The mutation may cause two splicing abnormalities:

- Exon 34 skipping.
- A 7-base deletion (non-triplet), leading to a frameshift and premature termination.

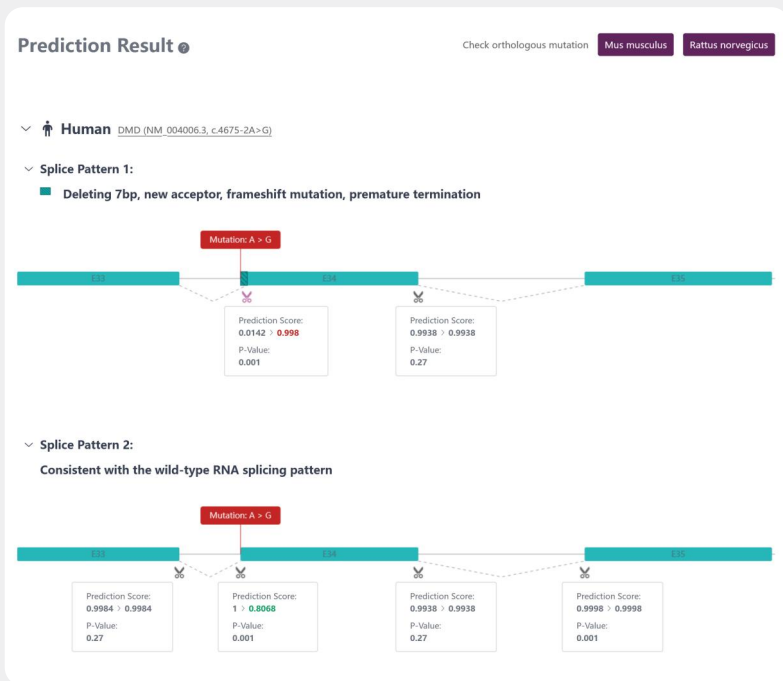


Figure 6. RNA Splicing Model Prediction

## ■ Validation Results

Literature reports confirm the same mutation in DMD carriers, generating a novel splice acceptor site (Hofstra et al., 2004), consistent with RDDC's predictions. To date, over 4,000 pathogenic DMD mutations have been reported. RNA Splicer's predictions for dozens of DMD splicing mutations show high concordance with published results.

# Empowering Rare Disease Research with AI-Driven Insights

---



## AI-Driven Insights

Leveraging vast datasets on diseases, genes, and mutations, RDDC employs machine learning and deep learning algorithms to develop tools for rare disease research and genetic disorder diagnosis.

---



## Core Data Focus

Centered on disease, gene, and animal model data, RDDC provides essential support for preclinical research and drug development in rare diseases.

---



## Data Visualization

With advanced data visualization and comprehensive comparison tools, RDDC enhances researchers' efficiency in information retrieval, addressing the shortcomings of traditional biological databases that prioritize data over user experience.

---



## Targeting Mutations

Focusing on the relationship between mutations and phenotypes, RDDC utilizes AI tools to explore the pathogenic mechanisms of genetic rare diseases, helping users pinpoint target genes and mutations with precision.

---



## Efficient AI Assistance

Integrating proprietary database APIs with large language models, RDDC allows users to query information conversationally, significantly reducing the learning curve and improving accessibility.

---

## Guangzhou Rare Disease Gene Therapy Alliance

Phone: 18028568424

Email: [rddc-support@tsinghua-gd.org](mailto:rddc-support@tsinghua-gd.org)

Website: <https://rddc.tsinghua-gd.org/>



Scan the QR code to access RDDC



Join the Alliance Today