



OPEN

DATA DESCRIPTOR

# RSDB: A rare skin disease database to link drugs with potential drug targets for rare skin diseases

Tien-Chueh Kuo<sup>1,2</sup>, Pei-Hua Wang<sup>2</sup>, Yu-Ke Wang<sup>2</sup>, Chia-I. Chang<sup>2</sup>, Ching-Yao Chang<sup>2</sup> & Yufeng Jane Tseng<sup>2,3</sup>✉

Rare skin diseases include more than 800 diseases affecting more than 6.8 million patients worldwide. However, only 100 drugs have been developed for treating rare skin diseases in the past 38 years. To investigate potential treatments through drug repurposing for rare skin diseases, it is necessary to have a well-organized database to link all known disease causes, mechanisms, and related information to accelerate the process. Drug repurposing provides less expensive and faster potential options to develop treatments for known diseases. In this work, we designed and constructed a rare skin disease database (RSDB) as a disease-centered information depository to facilitate repurposing drug candidates for rare skin diseases. We collected and integrated associated genes, chemicals, and phenotypes into a network connected by pairwise relationships between different components for rare skin diseases. The RSDB covers 891 rare skin diseases defined by the Orphanet and GARD databases. The organized network for each rare skin disease comprises associated genes, phenotypes, and chemicals with the corresponding connections. The RSDB is available at <https://rsdb.cmdm.tw>.

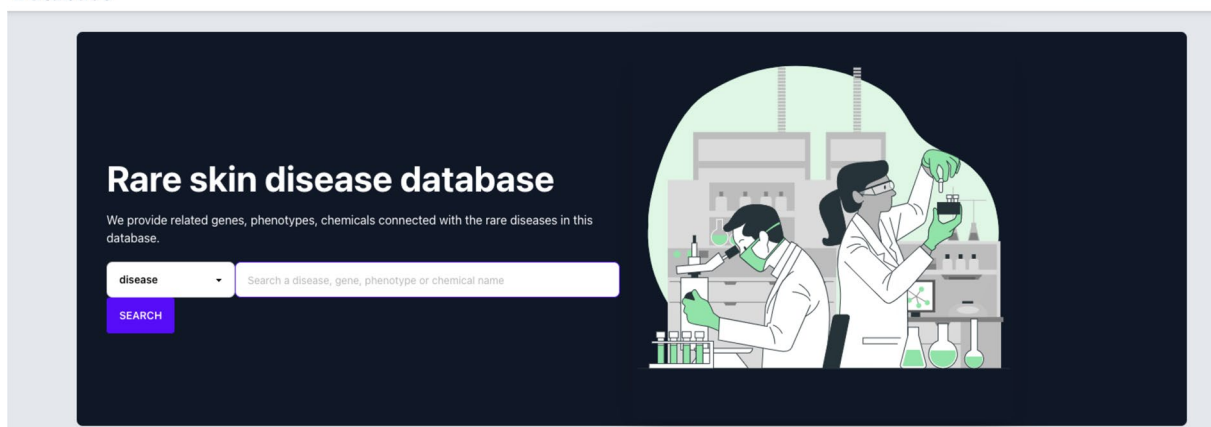
## Background & Summary

Rare diseases affect fewer than 1 in 200,000 people in the U.S. or 1 in 2,000 people in Europe<sup>1,2</sup>. Although most rare diseases are complex, disabling, and life-threatening<sup>3</sup>, they lack related studies and approved treatments<sup>4</sup> due to the limited prevalence and market<sup>5</sup>. Skin diseases cause significant nonfatal disability worldwide<sup>6</sup>, especially in resource-poor regions<sup>7</sup>. However, far little attention has been given to rare skin diseases<sup>8</sup>. In addition to the physiological burden, skin diseases' economic and social impacts significantly lower patients' quality of life<sup>9,10</sup>. Therefore, this work aims to help link drugs to drug targets for rare skin diseases.

Two databases, Orphanet and GARD, provide curated information on the diagnosis and currently available treatments for rare diseases<sup>11</sup>. Orphanet ([www.orpha.net](http://www.orpha.net)) covers rare diseases and orphan drugs, gathering and providing complete information and knowledge to improve diagnosis<sup>12</sup>. GARD, the Genetic and Rare Diseases Information Center, is a National Center for Advancing Translational Sciences (NCATS) program in the United States. It was established by the National Institutes of Health (N.I.H.) to provide information about symptoms, prevalence statistics, causes, treatments, diagnosis, and the latest research resources for over 6500 rare diseases<sup>13</sup>.

Although genetics accounts for the various causes of skin symptoms, the challenges of rare skin diseases are that they cannot be easily classified as skin disorders with a fixed set of symptoms. These symptoms vary from disease to disease and among patients with the same disease. Epidermolysis bullosa (E.B.) is a family of devastating rare skin diseases with friction inflicting painful, open wounds within the skin and internal epithelial tissue blistering<sup>14–16</sup>. Recent E.B. research has led to identifying mutations in 10 different genes<sup>17,18</sup>. One of the most severe forms of E.B. is recessive dystrophic epidermolysis bullosa (RDEB), caused by mutations in a protein called collagen VII<sup>19</sup>. Collagen VII provides the skin with structural integrity. There are over 500,000 people worldwide who suffer from this debilitating disorder. Simply looking at a single mutation or open wound

<sup>1</sup>The Metabolomics Core Laboratory, Center of Genomic Medicine, National Taiwan University, No. 1, Sec. 4, Roosevelt Road, Taipei, 10617, Taiwan. <sup>2</sup>Graduate Institute of Biomedical Electronics and Bioinformatics, College of Electrical Engineering and Computer Science, National Taiwan University, No. 1, Sec. 4, Roosevelt Road, Taipei, 10617, Taiwan. <sup>3</sup>Department of Computer Science and Information Engineering, College of Electrical Engineering and Computer Science, National Taiwan University, No. 1, Sec. 4, Roosevelt Road, Taipei, 10617, Taiwan. ✉e-mail: [yjtseng@csie.ntu.edu.tw](mailto:yjtseng@csie.ntu.edu.tw)



**Fig. 1** Search engine and homepage of the RSDB.

would not help identify the disease or the treatment. Expansion of the symptoms (phenotypes) to look for more probable treatment is needed.

Drug repurposing can reduce the risk of failure and the massive cost of money and time in drug development by identifying new indications for an existing drug that is already approved<sup>20,21</sup>. Drug repurposing aims to find new relationships between the drug and disease<sup>22</sup>. However, related data regarding rare skin diseases are scattered and stored in several biomedical databases. Most patient-centered databases provide diagnostic criteria or currently available treatments and prognoses. We collected and integrated associated genes, chemicals, and phenotypes into a network to find novel drug-disease relationships for rare skin diseases. The rare skin disease database (RSDB) covers 891 rare skin diseases defined by the Orphanet and GARD databases. The organized network for each rare skin disease comprises associated genes, phenotypes, and chemicals connected via associations found in PubChem<sup>23</sup>, MeSH<sup>24</sup>, the Comparative Toxicogenomics Database (CTD)<sup>25</sup>, and Human Phenotype Ontology (HPO)<sup>26</sup>. The RSDB is available at <https://rsdb.cmdm.tw>.

## Methods

We collected data from public databases containing curated, inferred, literature-based information to create a database for connecting biomedical information. With curated disease genes, phenotypes, and phenotype genes as the direct molecular signatures of rare skin diseases, this work tries to link potential drugs to candidate rare skin disease targets with matched genes through disease-gene or disease-phenotype-gene relationships.

Currently, the RSDB contains 891 rare skin diseases, 28,077 genes, 9,732 phenotypes and 17,297 compounds with 16,411 disease-gene relationships, 15,793 disease-phenotype relationships, 12,184 disease-reference relationships, 641,789 gene-phenotype relationships, 17,636 gene-reference relationships and 61,282 references. The RSDB will be updated twice a year in June and December.

Users can visit the RSDB homepage (<https://rsdb.cmdm.tw>) to explore the data for rare skin disease information. On the RSDB website, users can access records and perform searches (see Fig. 1).

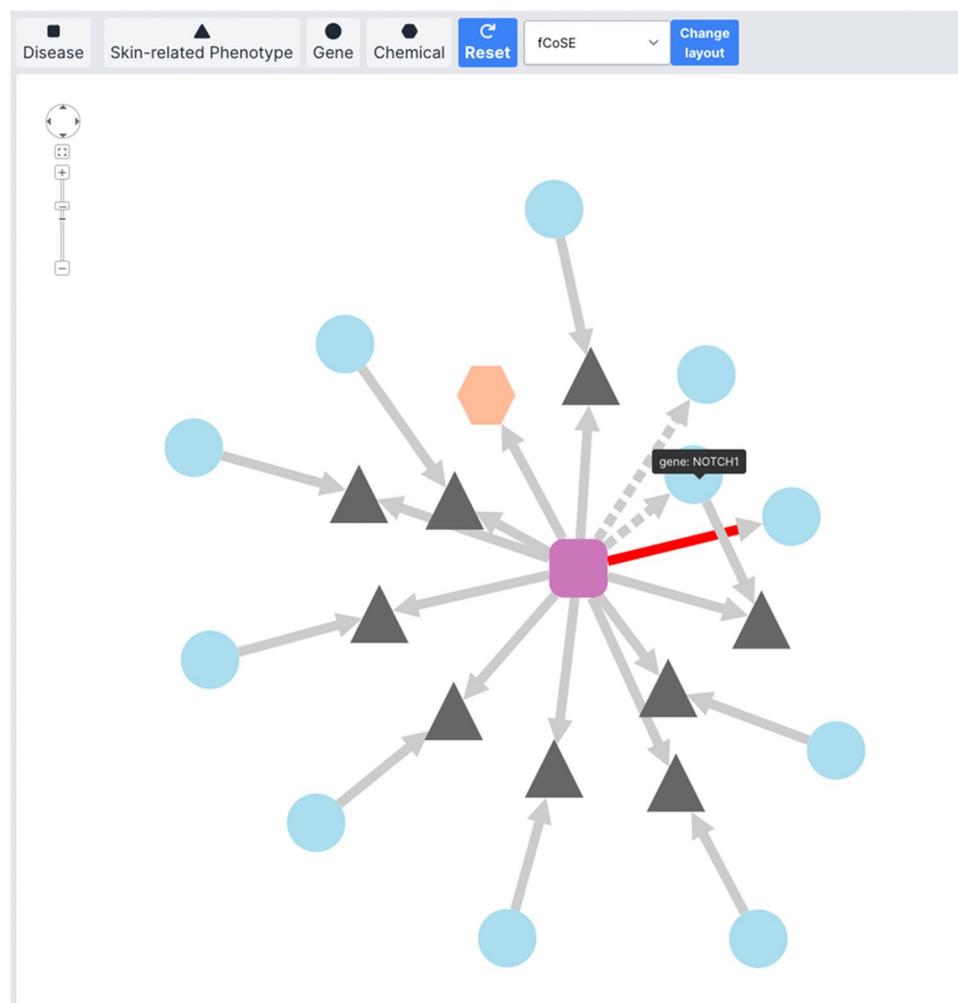
**Chemicals.** A total of 17,297 environmental chemicals including approved drugs, were imported from the dataset of chemicals to genes in the CTD and DrugBank. All chemicals associated with genes are included in the RSDB.

**Diseases.** Rare skin diseases were collected from Orphanet and GARD. Orphanet provides the disease classifications. All the rare diseases classified to the skin class were parsed and stored in the database.

The skin disease category was derived from NIH GARD. To determine whether a disease is a rare skin disease, we compared Orphanet as a basis. All information was downloaded, including the synonyms, definitions, inheritance, prevalence, and genes related to the disease. According to the NIH GARD, we found that 619 skin diseases have been described.

**Genes and disease-gene relationships.** Associated disease-gene relationships were collected from DisGeNET v7<sup>27</sup>. DisGeNET provides three tiers: (1) expert-curated information, (2) inferred information, and (3) text-mining information. Expertly curated information was collected from UniProt, the CTD, Orphanet, ClinGen, and Genomics England. (2) Inferred information was collected from NCBI ClinVar, HPO, the GWAS Catalog, and GWASdb<sup>28</sup>. (3) Text-mining information was collected from the LHGDN and BeFree system.

**Phenotypes and disease-phenotype and gene-phenotype relationships.** Associated phenotypes were collected from HPO and GARD. HPO provides disease-phenotype and gene-phenotype information. GARD provides rare disease-to-phenotype relationship information. We downloaded the 2020-12 version.



**Fig. 2** Screenshot of the network for cutis marmorata telangiectatica congenita (ORPHA: 1556). Nine phenotypes, two genes, and one chemical are directly connected to the disease, and eight genes are indirectly linked to the disease through phenotypes. The *NOTCH1* gene is related to the disease both directly and indirectly. Red edge stands for the curated disease-gene information.

**References.** Associated references were collected from the literature section of PubChem, which is linked to PubMed.

**Source database.** All data from different public databases were collected as follows.

**Expertly curated information.** UniProt<sup>29</sup>, the CTD<sup>30</sup>, Orphanet<sup>31</sup>, ClinGen<sup>32</sup>, Genomics England<sup>33</sup>

The CTD includes manually curated data on how chemicals interact with genes and proteins. Specifically, a chemical compound may interact with a gene or protein and influence its expression, folding, localization, activity, binding, abundance, and metabolic processing.

**Inferred information.** NCBI ClinVar<sup>34</sup>, HPO<sup>26</sup>, the Genome-Wide Association Study (GWAS) Catalog<sup>35</sup>.

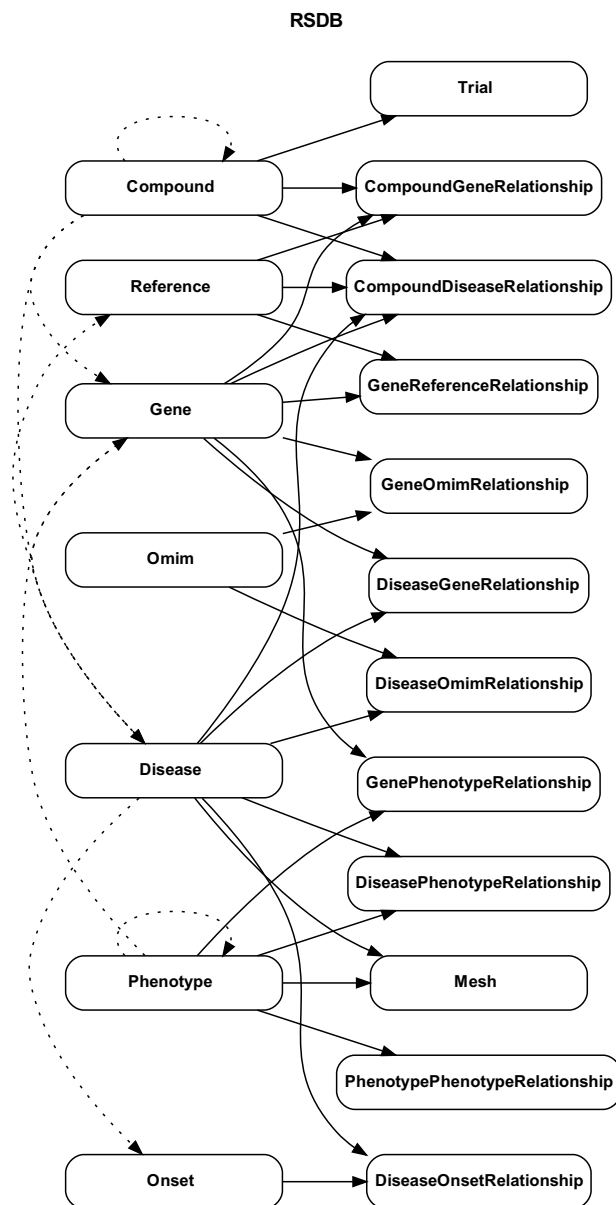
**Literature-based information.** The literature-derived human gene-disease network (LHGDN)<sup>36</sup>, BeFree system<sup>37</sup>.

### Data Records

All the data files in RSDB are stored in the Synapse repository (<https://doi.org/10.7303/syn34512708>)<sup>38</sup> and are available under the terms of CC BY-NC-SA 4.0 (<https://creativecommons.org/licenses/by-nc-sa/4.0/>).

There are 22 CSV files in the repository. Among them are nine files describing the basic components in RSDB, including compounds, genes, phenotypes, etc. The other 13 files store the pairwise relationships between components.

We designated an internal ID for all the files to every entry in the first column. For the files describing basic components, associated properties like names, descriptions, and ID numbers from other databases will be stored in the following columns. For the files describing relationships, we separate the many-to-many relationships in RSDB into multiple entries of pairwise relationships. For example, `disease_gene_relationships.csv` stores internal



**Fig. 3** Entity-relationship diagram of the RSDB.

disease ID and gene ID in the first and second columns, respectively. Disease with internal ID 3 is linked to the genes with internal ID 3 and 4 in the third and fourth entries. One can refer to `diseases.csv` and `genes.csv` for more information about the disease and genes involved in the relationships.

### Technical Validation

The datasets were retrieved from several public databases. According to the source database, the information we provide is curated by an expert or inferred from the literature or experiments. For example, our database connected four genes to the rare skin disease “exfoliative ichthyosis”: CSTA, KRT1, KRT2, and SERPINB8. Mutation in CSTA, which encodes cystatin A, can cause the disease<sup>39,40</sup>. Genetic linkages between the disease and KRT1 and KRT2, encoding keratin 1 and 2<sup>41</sup>, respectively. Loss-of-function mutations in SERPINB8, encoding serpin B8, are also linked to exfoliative ichthyosis<sup>42</sup>. The rare skin disease “epidermolytic palmoplantar keratoderma” has been confirmed to be caused by mutations in KRT1<sup>43</sup>, KRT9<sup>44</sup>, and KRT16<sup>45</sup>. This literature, which proves the accuracy of the disease-gene relationships in our data, is also provided to users via links to PubMed.

Here we demonstrate how our database can help drug repurposing using the well-known case of diacerein. Diacerein is a symptomatic drug in osteoarthritis. Its active metabolite, Rhein, decreases inflammation, reduces damage, and promotes the formation of new cartilage<sup>46</sup>. Diacerein has been effective against epidermolysis bullosa (EB) in the past decade by reducing blister counts and increasing skin stability<sup>47</sup>. There are four main types of EB, namely EB simplex (EBS), junctional EB (JEB), dystrophic EB (DEB), and Kindler syndrome (KS), according to the current international consensus classification. In RSDB, five genes directly link to chemical

diacerein: ACAN, COL1A1, COL2A1, and IL1B. Among them, COL1A1 and COL2A2 are linked to “dystrophic epidermolysis bullosa” (DEB) and “localized dystrophic epidermolysis bullosa, pretibial form,” a subtype of DEB, respectively. This validates our data and shows the possibility of finding a potential drug for repurposing.

The RSDB includes all the pairwise relationships between disease, gene, phenotype, and chemical-disease and chemical-gene associations. For a particular rare skin disease, the profile of the disease and lists of associated genes, phenotypes, or chemicals are provided along with network visualization. Integrated information that only multiple searches across several databases can obtain is organized into one webpage. Crosslinks to other databases and related articles in PubMed facilitate further analysis and study.

One outstanding feature of the RSDB is network visualization. Diseases, phenotypes, genes, and chemicals are denoted by pink squares, gray triangles, blue circles, and orange hexagons, respectively. For the network containing more than 50 nodes, the CiSE layout<sup>48</sup> will be applied to generate several circular layouts for each type of node to visualize the entire network without overlapping nodes. Otherwise, the fCoSE layout<sup>49</sup> will be applied. In addition, several layout algorithms, including circle, concentric, and CoSE layouts, are also available for users to change different network layouts. To access the node name and the link to the node page, users can click on the node, and the node information and link will appear in the tooltip. To pan, zoom in and out of the network, a navigation toolbar is provided on the top-left of the network. Network visualization helps users find genes and phenotypes relevant to particular rare skin diseases.

A gene can be indirectly linked to a disease in the network if both nodes are connected to the same phenotype, an intermediate node. For example, the gene “NOTCH1”, shown in Fig. 2, links to the disease directly and indirectly through a phenotype with HPO ID 25107. Multiple sources that lead to the same connection between one pair of diseases and genes imply a strong relationship between the disease and gene. We hope these findings help scientists find promising research targets and accelerate orphan drug discovery.

We developed a disease-centered database covering 891 rare skin diseases with associated genes, phenotypes, and chemicals. We deployed a complete text search engine to include exact matches and fuzzy searches for the search terms. On each chemical/disease/gene/phenotype page, all associated chemical/disease/gene/phenotype information is connected and visualized in the network. In the associated chemical/disease/gene/phenotype tables, all associated data will be listed with data source and evidence. The associated data can be filtered with keywords via the top-right search form of the tables.

## Usage Notes

**Overview of the RSDB.** We designed the RSDB with critical components, including (1) rare skin diseases, (2) genes, (3) phenotypes, and (4) chemicals. All four elements were collected from manually curated databases and connected with the associated information. All related information of one disease is seen as the molecular signature of the disease. An entity-relationship diagram is displayed in Fig. 3.

## Code availability

The code supporting this study’s findings is available on GitHub at [https://github.com/CMDM-Lab/rsdb\\_publication](https://github.com/CMDM-Lab/rsdb_publication).

The scripts and packages used for the RSDB rely on open-source packages such as Ruby on Rails, MariaDB, Elasticsearch, Cytoscape.js<sup>50</sup>, and in-house Ruby scripts.

Received: 27 January 2022; Accepted: 19 August 2022;

Published online: 26 August 2022

## References

1. Boycott, K. M., Vanstone, M. R., Bulman, D. E. & MacKenzie, A. E. Rare-disease genetics in the era of next-generation sequencing: discovery to translation. *Nat Rev Genet* **14**, 681–691, <https://doi.org/10.1038/nrg3555> (2013).
2. Simone Baldovino, M., Domenica Taruscio, M. & Dario Roccotello, M. Rare diseases in Europe: from a wide to a local perspective. *Sat* **12**, 20 (2016).
3. Navarrete-Opazo, A. A., Singh, M., Tisdale, A., Cutillo, C. M. & Garrison, S. R. Can you hear us now? The impact of health-care utilization by rare disease patients in the United States. *Genetics in Medicine* **23**, 2194–2201 (2021).
4. Hoeger, P. Genes and phenotypes in vascular malformations. *Clinical and Experimental Dermatology* **46**, 495–502 (2021).
5. Sardana, D. *et al.* Drug repositioning for orphan diseases. *Briefings in bioinformatics* **12**, 346–356 (2011).
6. Hay, R. J. *et al.* The global burden of skin disease in 2010: an analysis of the prevalence and impact of skin conditions. *Journal of Investigative Dermatology* **134**, 1527–1534 (2014).
7. Seth, D., Cheldize, K., Brown, D. & Freeman, E. E. Global burden of skin disease: inequities and innovations. *Current dermatology reports* **6**, 204–210 (2017).
8. McGrath, J. A. Rare inherited skin diseases and the Genomics England 100,000 Genome Project. *British Journal of Dermatology* **174**, 257–258 (2016).
9. Finlay, A. Y. The burden of skin disease: quality of life, economic aspects and social issues. *Clinical Medicine* **9**, 592 (2009).
10. Zhang, X.-j. *et al.* The psychosocial adaptation of patients with skin disease: a scoping review. *BMC public health* **19**, 1–15 (2019).
11. Shen, F. *et al.* Rare disease knowledge enrichment through a data-driven approach. *BMC medical informatics and decision making* **19**, 1–11 (2019).
12. Rath, A. *et al.* Representation of rare diseases in health information systems: the Orphanet approach to serve a wide range of end users. *Human mutation* **33**, 803–808 (2012).
13. Zhu, Q. *et al.* An integrative knowledge graph for rare diseases, derived from the Genetic and Rare Diseases Information Center (GARD). *Journal of Biomedical Semantics* **11**, 1–13 (2020).
14. Mellerio, J. E. *et al.* Emergency management in epidermolysis bullosa: consensus clinical recommendations from the European reference network for rare skin diseases. *Orphanet Journal of Rare Diseases* **15**, 1–10 (2020).
15. Montaudí, H., Chiaverini, C., Sbidian, E., Charlesworth, A. & Lacour, J. Inherited epidermolysis bullosa and squamous cell carcinoma: a systematic review of 117 cases. *Orphanet journal of rare diseases* **11**, 1–12 (2016).
16. Proding, C., Laimer, M., Bauer, J. W. & Hintner, H. EB (epidermolysis bullosa)-House Austria: Pioneering work for the care of patients with rare diseases. *JDDG: Journal der Deutschen Dermatologischen Gesellschaft* **18**, 1229–1235 (2020).



17. Chiu, F. C., Doolan, B., McGrath, J. & Onoufriadis, A. A decade of next-generation sequencing in genodermatoses: the impact on gene discovery and clinical diagnostics. *British Journal of Dermatology* **184**, 606–616 (2021).
18. Sawamura, D., Nakano, H. & Matsuzaki, Y. Overview of epidermolysis bullosa. *The Journal of dermatology* **37**, 214–219 (2010).
19. Wong, T. *et al.* Potential of fibroblast cell therapy for recessive dystrophic epidermolysis bullosa. *Journal of Investigative Dermatology* **128**, 2179–2189 (2008).
20. Oprea, T. & Mestres, J. Drug repurposing: far beyond new targets for old drugs. *The AAPS journal* **14**, 759–763 (2012).
21. Cha, Y. *et al.* Drug repurposing from the perspective of pharmaceutical companies. *British journal of pharmacology* **175**, 168–180 (2018).
22. Xue, H., Li, J., Xie, H. & Wang, Y. Review of drug repositioning approaches and resources. *International journal of biological sciences* **14**, 1232 (2018).
23. Kim, S. *et al.* PubChem in 2021: new data content and improved web interfaces. *Nucleic acids research* **49**, D1388–D1395 (2021).
24. Lipscomb, C. E. Medical subject headings (MeSH). *Bulletin of the Medical Library Association* **88**, 265 (2000).
25. Davis, A. P. *et al.* Comparative toxicogenomics database (CTD): update 2021. *Nucleic acids research* **49**, D1138–D1143 (2021).
26. Köhler, S. *et al.* The human phenotype ontology in 2021. *Nucleic acids research* **49**, D1207–D1217 (2021).
27. Piñero, J. *et al.* The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic acids research* **48**, D845–D855 (2020).
28. Li, M. J. *et al.* GWASdb: a database for human genetic variants identified by genome-wide association studies. *Nucleic acids research* **40**, D1047–D1054 (2012).
29. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research* **49**, D480–D489 (2021).
30. Mattingly, C. J., Colby, G. T., Forrest, J. N. & Boyer, J. L. The comparative toxicogenomics database (CTD). *Environmental health perspectives* **111**, 793–795 (2003).
31. Orphanet: an online rare disease and orphan drug data base. Copyright, INSERM. Available on <http://www.orpha.net>. Accessed 2021/04/28 (1999).
32. Rehm, H. L. *et al.* ClinGen—the clinical genome resource. *New England Journal of Medicine* **372**, 2235–2242 (2015).
33. The National Genomics Research and Healthcare Knowledgebase v5.1, Genomics England. (2020).
34. Landrum, M. J. *et al.* ClinVar: improvements to accessing data. *Nucleic acids research* **48**, D835–D844 (2020).
35. Buniello, A. *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic acids research* **47**, D1005–D1012 (2019).
36. Bundschuh, M., Dejori, M., Stetter, M., Tresp, V. & Kriegel, H.-P. Extraction of semantic biomedical relations from text using conditional random fields. *BMC bioinformatics* **9**, 1–14 (2008).
37. Bravo, A., Piñero, J., Queralt-Rosinach, N., Rautschka, M. & Furlong, L. I. Extraction of relations between genes and diseases from text and large-scale data analysis: implications for translational research. *BMC bioinformatics* **16**, 1–17 (2015).
38. Tien-Chueh, K. *et al.* Rare Skin Disease Database, *synapse*, <https://doi.org/10.7303/syn34512708> (2022).
39. Blaydon, D. C. *et al.* Mutations in CSTA, encoding Cystatin A, underlie exfoliative ichthyosis and reveal a role for this protease inhibitor in cell-cell adhesion. *The American Journal of Human Genetics* **89**, 564–571 (2011).
40. Moosbrugger-Martinez, V. *et al.* Epidermal barrier abnormalities in exfoliative ichthyosis with a novel homozygous loss-of-function mutation in CSTA. *British Journal of Dermatology* **172**, 1628–1632 (2015).
41. Steijlen, P. M. *et al.* Genetic linkage of the keratin type II gene cluster with ichthyosis bullosa of Siemens and with autosomal dominant ichthyosis exfoliativa. *Journal of investigative dermatology* **103**, 282–285 (1994).
42. Pigors, M. *et al.* Loss-of-function mutations in SERPINB8 linked to exfoliative ichthyosis with impaired mechanical stability of intercellular adhesions. *The American Journal of Human Genetics* **99**, 430–436 (2016).
43. Terron-Kwiatkowski, A. *et al.* Mutation S233L in the 1B domain of keratin 1 causes epidermolytic palmoplantar keratoderma with “tonotubular” keratin. *Journal of investigative dermatology* **126**, 607–613 (2006).
44. Rugg, E. *et al.* Diagnosis and confirmation of epidermolytic palmoplantar keratoderma by the identification of mutations in keratin 9 using denaturing high-performance liquid chromatography. *British Journal of Dermatology* **146**, 952–957 (2002).
45. Shamsher, M. *et al.* Novel mutations in keratin 16 gene underly focal non-epidermolytic palmoplantar keratoderma (NEPPK) in two families. *Human molecular genetics* **4**, 1875–1881 (1995).
46. Pavelka, K. *et al.* Diacerein: Benefits, Risks and Place in the Management of Osteoarthritis. An Opinion-Based Report from the ESCEO. *Drugs Aging* **33**, 75–85, <https://doi.org/10.1007/s40266-016-0347-4> (2016).
47. Bruckner-Tuderman, L. Newer Treatment Modalities in Epidermolysis Bullosa. *Indian Dermatol Online J* **10**, 244–250, [https://doi.org/10.4103/idoj.IDOJ\\_287\\_18](https://doi.org/10.4103/idoj.IDOJ_287_18) (2019).
48. Dogrusoz, U., Belviranlı, M. E. & Dilek, A. CiSE: A Circular Spring Embedder Layout Algorithm. *Ieee T Vis Comput Gr* **19**, 953–966, <https://doi.org/10.1109/Tvcg.2012.178> (2013).
49. Dogrusoz, U., Giral, E., Cetintas, A., Civril, A. & Demir, E. A layout algorithm for undirected compound graphs. *Inform Sciences* **179**, 980–994, <https://doi.org/10.1016/j.ins.2008.11.017> (2009).
50. Franz, M. *et al.* Cytoscape.js: a graph theory library for visualisation and analysis. *Bioinformatics* **32**, 309–311 (2016).

## Acknowledgements

We appreciate Dr. Yiumo Michael Chan for providing comments that improved the database. This work was supported by the Taiwan Ministry of Science and Technology (MOST 109-2823-8-002-010-CV, MOST 109-2320-B-002-040-, MOST 110-2320-B-002-038-); Taiwan Food and Drug Administration (MOHW110-FDA-D-114-000611); and National Taiwan University (NTU-CC-110L890803, NTU-110L8809). We thank the Laboratory of Computational Molecular Design and Metabolomics and the Department of Computer Science and Information Engineering of National Taiwan University for the resources used in performing these studies.

## Author contributions

Y.J.T. conceived the project. T.C.K., P.H.W., Y.K.W. and C.I.C. collected the data. T.C.K., P.H.W., Y.K.W., C.I.C. and Y.J.T. wrote the manuscript. T.C.K. designed and implemented the database. C.Y.C. implemented the database.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to Y.J.T.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022