Institution: Cardiff University

Unit of Assessment: Clinical Medicine (1)

Title of case study: The Human Gene Mutation Database: a major international resource for improved diagnosis of diseases and personalised genomics

Period when the underpinning research was undertaken: 2000 – 2020

Details of staff conducting the underpinning research from the submitting unit:

Name(s):	Role(s) (e.g. job title):	Period(s) employed by submitting HEI:
Cooper, David N.	Professor	01/11/1995 – present
Stenson, Peter	Research Associate	24/11/1997 – present
Mort, Matthew	Senior Software Developer	30/10/2000 – present
Hayden, Matthew	Research Assistant	04/03/2015 – present

Period when the claimed impact occurred: 1/8/2013 – 31/12/2020

Is this case study continued from a case study submitted in 2014? No

1. Summary of the impact (indicative maximum 100 words)

The Human Gene Mutation Database (HGMD), developed at Cardiff, is the first and only fully curated, annotated collection of inherited disease-causing mutations in nuclear genes. Cardiff licenced the HGMD to commercial partner QIAGEN, and it has now become the primary disease-associated mutation database used by the international biomedical and clinical community. This resource is employed worldwide by over 700 organisations in public health and commercial settings, in both clinical diagnostics and personalized genomics ensuring rapid and accurate test results.

2. Underpinning research (indicative maximum 500 words)

The Human Gene Mutation Database (HGMD) was first established in 1996 and consists of pathologically relevant mutation data from the peer-reviewed literature, curated and annotated for users by the Cardiff team. Since 2000, the HGMD has been developed into a valuable scientific resource, as follows:

In 2003, the HGMD was expanded to include cDNA reference sequences for more than 87% of listed genes, as well as splice junction sequences, and disease-associated and functional polymorphisms, vastly expanding its utility for clinical and scientific purposes. New entries were added at a rate of 5,000 per annum during this period **[3.1]**.

Following a further increase in data added to the database from 2005, it became the central unified repository for disease-related genetic variation in the germline. By 2008, new entries added by the Cardiff team exceeded 9,000 per annum. This enhanced the usability of the database beyond academia, specifically for human molecular geneticists, genome scientists, molecular biologists, clinicians, and genetic counsellors, as well as researchers specializing in biopharmaceuticals, bioinformatics and personalised genomics **[3.2]**.

The value of the HGMD was illustrated by a collaborative study involving the Cardiff team and US collaborators. The project demonstrated that individuals unselected for disease harboured disease-causing mutations associated with a phenotype either in themselves or their family more frequently than had been previously assumed **[3.3]**, with similar results found during a collaboration with the 1,000 Genomes Pilot Project **[3.4]**.

The HGMD also facilitated development of the computational model MutPred at Cardiff which predicts changes in protein sequences as a consequence of genetic mutation. First published in 2009, MutPred models changes in structural features and functional sites between wild-type and missense mutant sequences. MutPred was independently confirmed to be one of the best performing mutation pathogenicity prediction methods available (Thusberg et al., 2011 doi: 10.1002/humu.21445). A distinguishing feature of MutPred is the probabilistic modelling of variant impact on specific aspects of protein structure and function that can serve to guide experimental studies of phenotype-altering variants. The latest version, MutPred2, was shown to identify

structural and functional mutational signatures relevant to Mendelian disorders and the prioritization of de novo mutations associated with complex neurodevelopmental disorders **[3.5]**.

To further facilitate HGMD's utility for clinical groups and bioinformaticians, HGMD data formats were adapted to integrate with next-generation sequencing (NGS). NGS results can be directly compared with HGMD data, with relevant variants previously implicated in disease causation highlighted. This advance improved the usability of HGMD data and greatly increased the computational analyses that could be applied, enhancing application in clinical settings.

By the end of 2020, HGMD contained over 307,000 manually curated mutation reports in more than 12,000 genes, published in over 3,000 peer-reviewed journals: ~30,000 mutation entries are currently being added to the database per annum, and cDNA reference sequences are now available for 98% of listed genes **[3.6]**.

3. References to the research (indicative maximum of six references)

[3.1] Stenson PD, Ball EV, **Mort M**, Phillips AD, Shiel JA, Thomas NST, Abeysinghe A, Krawczak M, **Cooper DN**. (2003) Human Gene Mutation Database (HGMD): 2003 update. *Human Mutation* 21(6): 577-581. https://doi.org/10.1002/humu.10212

[3.2] Stenson PD, **Mort M**, Ball EV, Howells K, Phillips AD, Thomas NST, **Cooper DN**. (2009) The Human Gene Mutation Database (HGMD): 2008 update. *Genome Medicine* 1(1): article number 13. https://doi.org/10.1186/gm13

[3.3] Johnston JJ, Lewis KL, Ng D, Singh LN, Wynter J, Brewer C, Brooks BP, Brownell I, Candotti F, Gonsalves SG, Hart SP, Kong HH, Rother KI, Sokolic R, Solomon BD, Zein W, **Cooper DN**, **Stenson PD**, Mullikin JC, Biesecker LG. (2015) Individualized iterative phenotyping for genome-wide analysis of loss-of-function mutations. *Am J Hum Genet* 96(6): 913-925. https://doi.org/10.1016/j.ajhg.2015.04.013

[3.4] Xue Y, Chen Y, Ayub Q, Huang N, Ball EV, **Mort M**, Phillips AD, Shaw K, **Stenson PD**, **Cooper DN**, Tyler-Smith C, and the 1000 Genomes Pilot Project Consortium. (2012) Deleterious and disease allele prevalence in healthy individuals: insights from current predictions, mutation databases and population-scale resequencing. *Am. J. Hum. Genet* 91: 1022-1032. https://doi.org/10.1016/j.ajhg.2012.10.015

[3.5] Li B, Krishnan VG, **Mort ME**, Xin F, Kamati KK, **Cooper DN**, Mooney SD, Radivojac P. (2009) Automated inference of molecular mechanisms of disease from amino acid substitutions. *Bioinformatics* 25: 2744-2750. https://doi.org/10.1093/bioinformatics/btp528

[3.6] Stenson PD, **Mort M**, Ball EV, Chapman M, Evans K, Azevedo L, **Hayden M**, Heywood S, Millar DS, Phillips AD, **Cooper DN**. (2020) The Human Gene Mutation Database (HGMD®): optimizing its use in a clinical diagnostic or research setting. *Hum Genet*. 139:1197-1207. https://doi.org/10.1007/s00439-020-02199-3

4. Details of the impact (indicative maximum 750 words)

HGMD is the sector-leading database of disease-causing human gene variations. Several other centralised databases are available that record pathogenic variation, such as Online Mendelian Inheritance in Man (OMIM), ClinVar, dbSNP, and LOVD. The closest competitor, ClinVar, lacks depth compared to HGMD (in terms of variant and literature coverage) and obtains the majority of its pathogenic variant data via direct submission from clinical testing laboratories, thereby limiting its submission base. HGMD is the only database of pathological variants that approaches comprehensive coverage of the peer-reviewed literature, where there is evidence of clinical impact.

4.1 Commercial Market Position

Registered academic users (~150,000) can access a public version of HGMD for free (http://www.hgmd.org), with a subscription version (HGMD Professional) distributed through QIAGEN GmbH via a Licence Agreement with Cardiff University (distribution was initially through BIOBASE, until QIAGEN GmbH's acquisition of BIOBASE in May 2014) **[5.1]**.

QIAGEN currently has more than 600 HGMD Professional subscribers across 51 countries worldwide. Dr Frank Schacherer, VP Products and Solutions at QIAGEN, confirmed that the customer base has continued to grow, and the compound annual growth rate of sales over the last five years was 18%, "*thanks to the University's commitment to keep the database current, comprehensive and competitive*" **[5.2]**.

In July 2014, QIAGEN signed an agreement with BGI Tech Solutions Ltd, a subsidiary of BGI Group, the largest genomics organisation in the world. Under the terms of the agreement, BGI now provide HGMD data to the Greater China (Mainland China, Hong Kong, Macau, Taiwan) market as a distributor and provide first level support for the database within this market **[5.3]**.

4.2 Use of HGMD by clients

Both HGMD and HGMD Professional have a wide variety of uses, and as a result, a broad range of end-users worldwide **[5.1]**. In 2016, surveyed users reported that they used the database to save time trawling through literature, to determine if an identified mutation had been previously published, and to quickly find known mutations for a specific gene, solving challenges such as workflow bottlenecks, delayed processing of patient samples, and prioritising correct mutations **[5.4]**.

Schacherer stated: "*As the global use of genomics in research and development has increased, the Cardiff team has added content and functionality to HGMD to maintain its relevance and increase its utility*" **[5.2]**.

a. Private clients

Subscribers who use HGMD Professional to accelerate their diagnostic approaches, include John Hopkins University, Baylor College of Medicine, Mayo Clinic, MD Anderson Cancer Center, Scripps Research Institute, and the Children's Hospital of Philadelphia. Feedback received by QIAGEN on the use of HGMD by clients includes:

- Dr Ali Torkamani, CSO Human Longevity Inc. (USA): "*HGMD Professional provides the most comprehensive database of human disease associations and is an invaluable resource in both clinical and research-grade genetics and genomics activities*" **[5.2]**.
- Dr Yaping Yang, Co-Director of the Baylor College of Medicine DNA Diagnostic Laboratory: "*We rely on HGMD® professional heavily for reporting our clinical tests*" **[5.5]**.
- Anonymous lab director at a health care company: "*It has provided up-to-date information about genes and mutations that help facilitate the interpretation of patient results. It is an excellent tool and saves me a lot of time*" **[5.4]**.

Schacherer also stated that "*I had one customer who operates a genetic testing laboratory tell me outright that they could not do their work without HGMD*" **[5.2]**.

LabCorp, one of the largest clinical laboratory networks in the world, incorporated HGMD within their QIAGEN Clinical Insight licence in 2019 in order to improve identification and interpretation of genetic variants within inherited diseases. Marcia Eisenberg, Chief Scientific Officer for LabCorp Diagnostics, stated: "*Having access to the most comprehensive and up-to-date catalog of known mutations augments our existing variant classification expertise. This will allow us to continue to provide physicians and researchers with the best possible test interpretations, advancing LabCorp's mission to improve health and improve lives*" **[5.6]**.

b. Public Clients

HGMD data are used across the UK for diagnosis and prioritisation of candidate genes for disease analysis in a variety of NHS hospitals and foundations, including the Royal Brompton, Addenbrooke's, and Great Ormond Street. **[5.7]**

At a national level, Genomics England, a company set up and owned by the Department of Health and Social Care, uses HGMD Professional to deliver the 100,000 Genomes Project **[5.1]**. This project has sequenced 100,000 whole genomes from NHS patients with rare diseases (~17,000 patients) and common cancers (~25,000 patients), as well as their families and a control group of volunteers with no known genetic condition **[5.8]**.

The Director of Bioinformatics and Genomics England stated: "*Our experience of using QIAGEN's HGMD for the 100,000 Genomes Project guided our decision to continue to rely upon this industry-leading resource …it is critical for healthcare providers to be able to interpret NGS data in the context of the vast body of knowledge from research and clinical experience. The exhaustive knowledge in HGMD enables us to do this, by providing the best possible care*" **[5.9]**.

During sequencing of genomes in the 100,000 Genomes Project, many examples of HGMD-listed mutations were noted, even for project donors from the general population who appeared healthy, in line with Cardiff's research **[3.3, 3.4]**. These results were fed back to these participants via their hospital teams. As of March 2019, potential diagnoses were identified for approximately 3,300 (1 in 5) of the rare disease sufferers enrolled in the programme and for around 40% of those with intellectual disabilities. Clinical trials or more effective medicines were also identified for around half of the cancer patients enrolled (~12,500 patients), enhancing their care and increasing the chances of survival **[5.8]**.

4.3 Summary

Cardiff's HGMD has become a vital world-leading resource for a wide range of stakeholders (hospitals, companies, universities and governments) across the world, providing rapid access to up-to-date gene mutation data accelerating clinical research. This is evidenced by a significant increase in sales of HGMD Professional, diversifying and growing QIAGEN's customer base, including provision of HGMD within Greater China for the first time.

5. Sources to corroborate the impact (indicative maximum of 10 references)

[5.1] HGMD Professional Webpage

[5.2] Testimonial from Dr Frank Schacherer, VP Products and Solutions, QIAGEN GmbH

[5.3] QIAGEN press release: QIAGEN agreement to provide HGMD data throughout Greater China through BGI Tech (July 2014)

[5.4] QIAGEN 2016 Survey on use of HGMD, Qiagen paper

[5.5] QIAGEN HGMD information sheet

[5.6] Qiagen press release: LabCorp acquire rights to use HGMD (November 2019)

[5.7] Examples of uses of HGMD at NHS hospitals

[5.8] 100,000 Genomes Project details, Genomics England webpage

[5.9] Qiagen press release: Genomics England renew contract with QIAGEN to use HGMD, (February 2019)