Biophysical Reviews
https://doi.org/10.1007/s12551-022-00933-x

REVIEW

A bibliometric review of peripartum cardiomyopathy compared to other cardiomyopathies using artificial intelligence and machine learning

M. Grosser[1] · H. Lin[1] · M. Wu[2] · Y. Zhang[2] · S. Tipper[1] · D. Venter[1] · J. Lu[2] · C. G. dos Remedios[3,4]

Received: 21 December 2021 / Accepted: 24 January 2022

Abstract

As developments in artificial intelligence and machine learning become more widespread in healthcare, their potential to transform clinical outcomes also increases. Peripartum cardiomyopathy is a rare and poorly-characterised condition that presents as heart failure in the last trimester prior to delivery or within 5–6 months postpartum. The lack of a definitive understanding of the molecular causes and clinical progress of this condition suggests that bibliometrics will be well-suited to creating new insights into this serious clinical problem. We examine similarities and differences between peripartum and its closely related familial dilated cardiomyopathy and idiopathic dilated cardiomyopathy. Using PubMed as the source of bibliometric data, we apply artificial intelligence–supported natural language processing to compare extracted data and genes association with these cardiomyopathies. Gene data were enhanced with additional metadata from third-party datasets and then analysed for their impact and specificity for peripartum cardiomyopathy. Artificial intelligence identified 14 genes that distinguished peripartum from both dilated and familial dilated cardiomyopathy. They are as follows: **CTSD**, RLN2, MMP23B*, SLC17A5, ST2*, PTHLH, CFH*, CFI, **GPT**, MR1, Rln1, **SRI**, STAT5A* and THBD. We then used the Human Protein Atlas website that uses affinity-purified rabbit polyclonal antibodies to identify genes that are expressed at the protein level (bold), or as RNA transcripts (*) in healthy human left ventricles. Additional analysis focussed on the full set of peripartum genes on linkage and specificity to cardiomyopathy yielded a different set of thirteen genes (bold font indicates those expressed in cardiomyocytes: PRL, RLN2, **PLN**, ST2, **CTSD**, F2, ACE, STAT3, **TTN**, SPP1, LGALS3, miR-146a, **GNB3**, SRI). This type of analysis can highlight new avenues for research, aimed at improving genomics-driven peripartum cardiomyopathy diagnosis as well as potential pathological and clinical sub-classification. We expect that this will allow for future improvements in identification, treatment and management of this condition. The first step in the application of these bibliometric-based artificial intelligence methods is to understand the current knowledge, and it is the aim of this paper to show how this might be achieved.

Keywords Artificial intelligence · Genomics · Bibliometrics · Machine learning · Peripartum cardiomyopathy

Introduction

This study focuses on the use of novel computer-based approaches, such as artificial intelligence (AI) and machine learning (ML), to understanding the current landscape of diseases such as peripartum cardiomyopathy (PPCM). Our aim is to provide personal and precise insights for patients and their healthcare providers to identify causes, targeted treatments, improved management and outcomes based on a patient's genome.

The computer-aided analysis in this review has been done as a bibliometric study, which uses quantitative methods to describe characteristics of publications to assess the current

✉ M. Grosser
mark.grosser@23strands.com

1 23 Strands Pty Ltd, 107, 26 Pirrama Rd, Pyrmont, NSW, Australia

2 University Technology Sydney, 15 Broadway, Ultimo, NSW, Australia

3 Victor Chang Cardiac Research Institute, 405 Liverpool St, Darlinghurst, Australia

4 Sydney Heart Bank, University of Sydney, Sydney, Australia

Published online: 09 February 2022

state of a field and provide insight into its overall structure (Okubo 1997). The statistical techniques have been further refined and the methodology also includes enhanced AI based topic extraction, disambiguation and network and feature analysis (Wu et al. 2021).

The clinical focus of this study is peripartum cardiomyopathy (PPCM—also known as postpartum cardiomyopathy). PPCM is a rare form of heart failure that occurs during the last month of pregnancy or within 6 months of delivery (Elliott et al. 2007). A common hypothesis is that PPCM is caused by a combination of hemodynamic and metabolic stress (Bello and Arany 2015), leading to myocardial damage where the underlying triggers that cause the onset of this disease are still not established.

PPCM falls in the diagnostic class of cardiomyopathies which are a major cause of the cardiovascular burden on the health system. A recent and exhaustive study in France (Lannou et al. 2020) indicated that, in patients under the age of 40 years, cardiomyopathies accounted for 71% of all heart transplants, 51% of all cardiac assist devices, 63% of patients requiring circulatory support and 23% of heart failures.

Dilated cardiomyopathy (DCM), familial DCM (FDCM) and PPCM are often linked in the literature and often share clinical signs (ventricular dilation, impaired systolic function and arrhythmias), but there are clear reasons why PPCM should be considered a separate class of disease (Bollen et al. 2015). PPCM typically presents in a younger cohort of patients, with signs of heart failure commonly not recognised until the last month of pregnancy or more commonly within 6 months of delivery (Fig. 1), in the absence of other causes.

Apart from being a relatively rare, there are also differences in prevalence across different ethnic groups. The global incidence of PPCM (Sliwa et al. 2017; Isogai and Kamiya 2019) is significantly increased in women of African descent (Davis et al. 2020). PPCM is also often under-reported because its symptoms resemble the normal course of pregnancy (Lewey et al. 2020). These differences and the link to the third trimester of pregnancy led to possible familial causes of PPCM, and it is possible that the whole class of PPCM or one or more potential subtypes may fulfil the criteria of Mendelian disease.

DCM is defined as left and/or right ventricular dilation and cardiac muscle dysfunction without abnormal loading conditions (such as hypertension or valve disease), or coronary artery disease, sufficient to impair global heart function (McKenna et al. 2017). DCM is often cited as the most prevalent cardiomyopathy (McKenna and Judge 2021; Charron et al. 2018). A recent study from Sub-Saharan Africa (Fundikira et al. 2021) identifies it as the most prevalent in that community (incidence of 1:2,500), compared to the results in Europe (Lannou et al. 2020) where DCM was second to hypertrophic cardiomyopathy (HCM).

FDCM is clinically defined by the presence of DCM in an index patient who shares the disease with family members. Elliott et al. (2007) modified the DCM classification to include FDCM defined by the presence of mutations in the following functional gene categories: (1) sarcomeric genes; (2) sarcoplasmic genes (muscle LIMP Cytoskeletal genes, dystrophin, desmin, metavinculin, sarcoglycan complex, CRYAB, epicardin); (3) nuclear membrane genes (lamin A/C, emerin); (4) intercalated disc gene mutations; and (5) mitochondrial cytopathies.

Conventional approaches to identifying mutated genes that form the current scientific understanding of disease and their corresponding proteins are typically identified by inserting keywords (linked by "AND") into a PubMed search. Instead, we will use AI and ML to identify mutated genes and their corresponding proteins that are potentially involved PPCM, as well as its relationship to selected other cardiomyopathies.

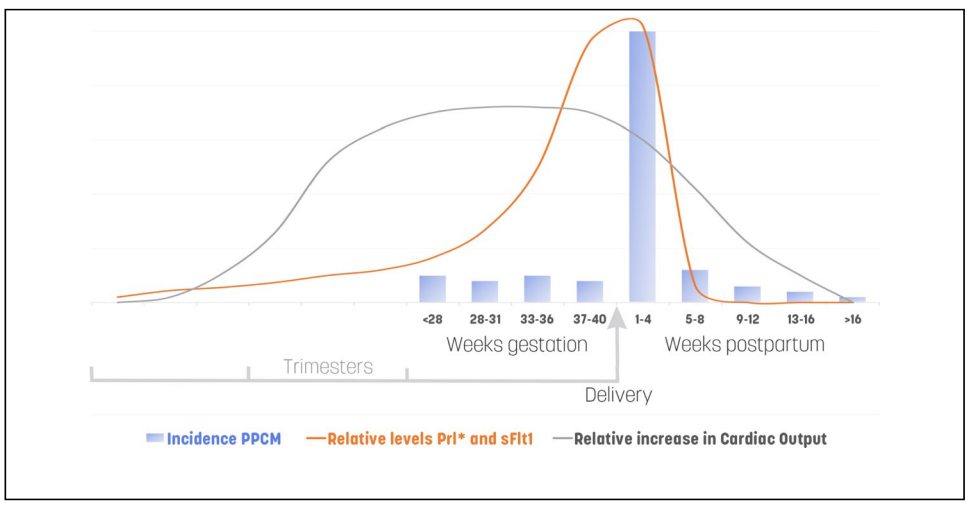

Fig. 1 The human pregnancy cycle with PPCM incidence where the columns represent the incidence; orange illustrates the circulating levels of prolactin (Prl) and soluble fms-like tyrosine kinase-1 (sFlt-1) during a normal pregnancy (Arany and Elkayam 2016)

The most recent definition by the United States National Library of Medicine (US NLM) lists cardiomyopathy (CM) as a group of diseases in which the dominant feature is the involvement of disease in the cardiac muscle, impacting the function of the heart. The library provides a clear hierarchy of classification of these diseases, shown in Table 1.

Here, we review the potential use of an AI-empowered assessment of the PPCM literature in the context of DCM and FDCM. We focus on their genetic aspects of the literature to see whether the disorder can be identified through the analysis of the genetic and genomic literature to improve our understanding of the predisposition, pathogenesis and potential therapeutic targets for this disorder. Specifically, we use intelligent bibliometrics and computational models and elaborating AI and data science techniques with bibliometric indicators (e.g. citation counts, contents and authorships) (Zhang et al. 2020).

Table 1 Hierarchy of cardiomyopathies in the US NLM (https://www.ncbi.nlm.nih.gov/mesh/68009202)

Cardiovascular Diseases
Heart Diseases
Cardiomyopathies
Arrhythmogenic Right Ventricular Dysplasia
Cardiomyopathy, Alcoholic
Cardiomyopathy, Dilated
Cardiomyopathy, Hypertrophic
Cardiomyopathy, Hypertrophic, Familial
Cardiomyopathy, Restrictive
Chagas Cardiomyopathy
Diabetic Cardiomyopathies
Endocardial Fibroelastosis
Isolated Noncompaction of the Ventricular Myocardium
Endomyocardial Fibrosis
Glycogen Storage Disease Type IIb
Kearns-Sayre Syndrome
Myocardial Reperfusion Injury
Myocarditis
Sarcoglycanopathies

The absence of FDCM and PPCM in the NLM classification suggests that improvements are required in the classification of journal articles to account for the emergence of newly recognised clinically important diseases, particularly those that are rare. The current lack of classification of these two categories (Elliott et al. 2007) increases the risk that the absence of structured classification may obscure "signals" in the exponentially growing scientific literature, particularly the increasingly complex understanding of digital searching. This lack of classification may also obscure analysis and identification of new insights into the underlying biology, when seeking improved diagnostics, treatments and management of patients with these diseases.

We will illustrate how bibliometrics, empowered by AI based methods, can assist researchers to better understand and potentially differential DCM, FDCM and PPCM.

Our focus is to better understand the molecular causes of PPCM and how they can lead to pathways of healthy cellular function compared with potential disease pathways. Figure 2 highlights the main locations of main contractile proteins and where mutations of some genes may be important in PPCM.

The genes relating to the list of proteins shown in Fig. 2 include the following: the super family of myosin heavy chains (MYH6, MYH7, MYH7B), myosin light chains (MYL2, MYL7), actin (ACTC1, ACTN2), myosin binding protein C (cMyBP-C), cardiac troponins (TNNT2, TNNC1, TNNI3), titin (TTN) and cardiac tropomyosin (TPM1).

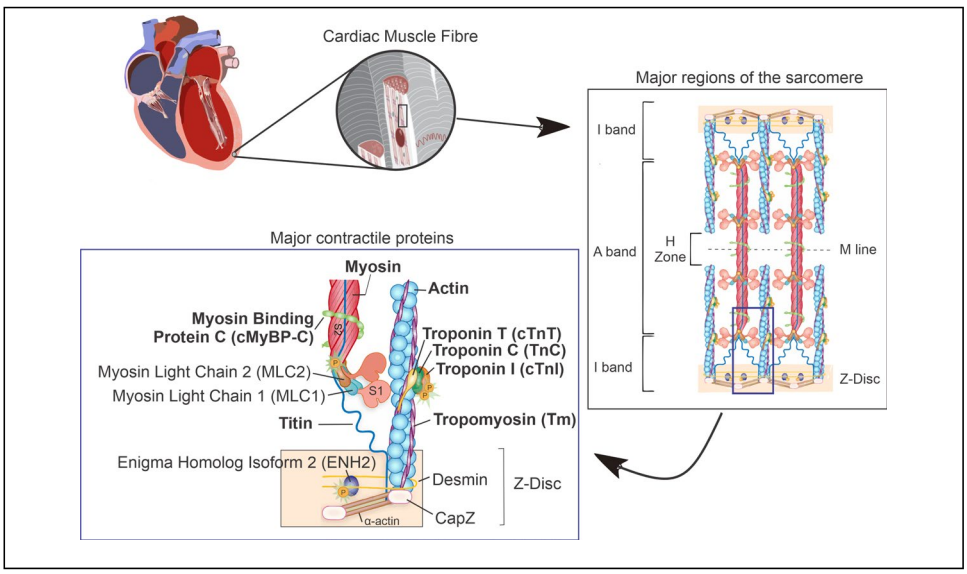

Fig. 2 Schematic summary of the key cardiac contractile proteins (adapted from Peng et al. 2014)

These genes were generated by extracting the related genes linked to the heart proteins from the Human Protein Atlas (Thul and Lindskog 2018) Table 2.

With the aid of intelligent bibliometrics, we address the following research questions:

- Which genes are most commonly mutated in PPCM and how do they compare with DCM, FDCM and unspecified cardiomyopathy?
- What are the major institutions and who are the authors of PPCM research?
- How can AI and ML review of other cardiomyopathies to understanding PPCM?
- What are the challenges that arise when reviewing a limited researched space?
- Based on the potential genes involved, what future opportunities arise for research directed towards improved diagnosis and therapy of PPCM?

Methods and materials

The bibliometric review framework of this study is given in Fig. 3.

Data collection

Falagas et al. (2008) recommend PubMed as the optimal bibliometric database for medical and life sciences. It enables biomedical information retrieval for use in specialized meta-classification data like Medical Subject Heading (MeSH) search and biomedical filters (including the humans filter we used in this work) to return precise results.

A naïve search strategy was used on title and abstracts which is a common screening approach in bibliometrics. We use the "gene*" (where * captures all terms commencing with "gene") criterion; we gather articles with high specificity to the genomics focus of this research. The fha[filter] and human[filter] ensure a focus on human-based research that has an abstract (fha).

Figure 4 illustrates that bibliometrics review is growing rapidly. While this approach has strengths and weaknesses (Penning de Vries et al. 2020; Linnenluecke et al. 2020), we intend to show that a computer-aided bibliometric approach is a good first step in understanding the research landscape on a topic.

The additional AI methods in bibliometrics empower cleaning, filtering and sentiment algorithms and allow easier extraction of higher quality topic extraction and structural information, such as gene information, disease

Table 2 Sarcomeric proteins identified in the literature being involved in DCM

SARCOMERIC	Body part	Location	Gene	T = trunc: M = missense	Disease	#	Impact on	Action	Reference
BCL2 Associathogene 3	Cardiac	Z disc	BAG3	T	PPCM, FDCM	1	Hsp70, small Hsps	Co-chaperone, autophagy	Goli et al. 2021
Myomesin 1	Skeletal	M line	MYOM1	pE247K	PPCM, FDCM	1	Obscurin, titin	Stretch sensor	Marston et al. 2015
Myosin binding protein C3	Skeletal	C zone within A band	MYBPC3	M	PPCM, FDCM, HCM		Actin, myosin, titin	Regulates actomyosin force	Morales et al. 2010
Myosin heavy chain 6	Skeletal	A band	MYH6	T, M	PPCM, HCM, AF	2	Actin, titin	Motor protein	Ware et al. 2016
Myosin heavy chain 7 beta	Skeletal	A band	MYH7	M, nN16499K	PPCM, FDCM, HCM	2	Actin, titin	Motor protein	Ware et al. 2016
Obscurin	Skeletal	A band	OBSCN	M, pV2161D	PPCM		Titin	Thick fil/myofibrillogenesis	Marston et al. 2015
Troponin C1	Skeletal	I band	TNNC1	M	PPCM, DCM		TNNI3, actin	Ca regulation	van Spaendonck-Zwarts et al. 2010
Troponin I3	Cardiac only	I band	TNNI3	M, pK36Q	PPCM, FDCM		TNNC1	Ca regulation	Marston et al. 2015
Troponin T2	Cardiac	I band	TNNT2	M	PPCM, HCM		Tropomyosin	Ca regulation	Spracklen et al. 2021a, b
Tropomyosin	Cardiac	I band	TPM	T	PPCM, FDCM		Actin, MYBPC3, TNNT	Strong loss-of-function	Spracklen et al. 2021a, b

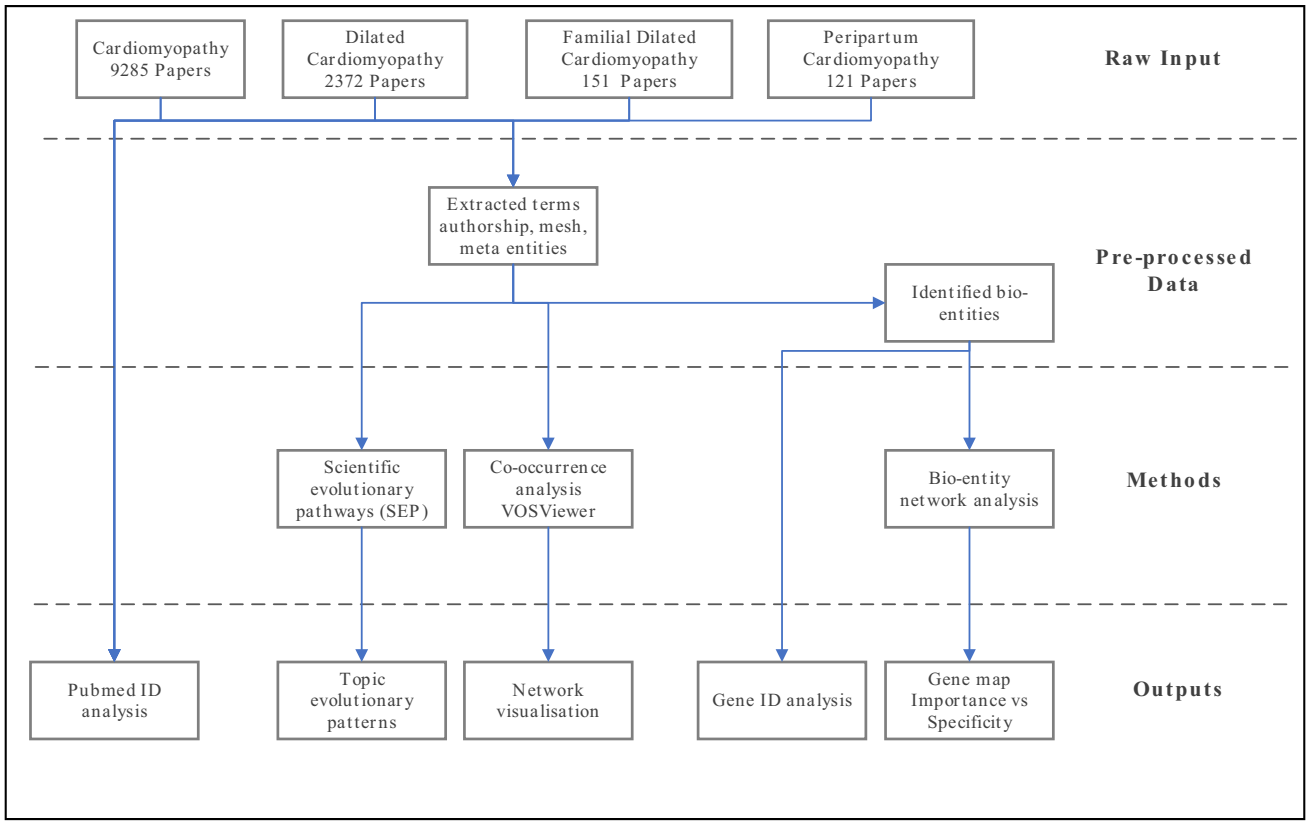

Fig. 3 Review framework showing the process steps used in this review

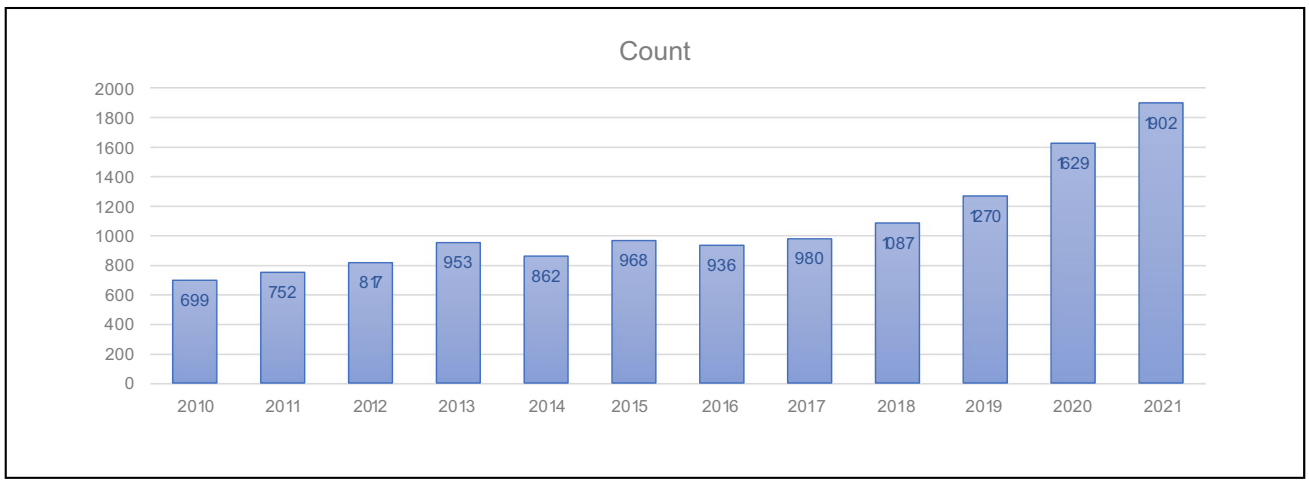

Fig. 4 The number of articles per year obtained from on Bibliometrics in PubMed indicates that this is a rapidly growing field of research

classification and chemical interactions. Using the specific naïve search strategies below, we assembled a dataset of the gene-related research performed on CM, DCM, FDCM and PPCM from PubMed, between 2005 and 2022. We included cardiomyopathy as a broad category to give a scale to the total research landscape universe:

CM Search query: **((**"**Cardiomyopathy**"**[Title/Abstract] AND** "**gene***"**[Title/Abstract]) AND ((fha[Filter]) AND (humans[Filter]))** Yielded: 9285 papers.

DCM Search query: **((**"**Dilated cardiomyopathy**"**[Title/Abstract] AND** "**gene***"**[Title/Abstract]) AND**

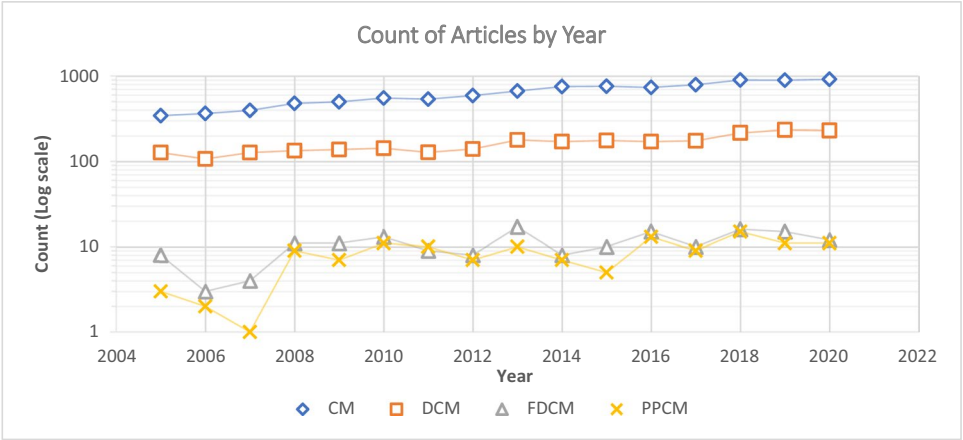

Fig. 5 The timeline of journal articles published in PubMed (note the log scale)

((fha[Filter]) AND (humans[Filter])) Yielded: 2372 Papers.

FDCM Search query: **(("Familial Dilated cardiomyopathy"[Title/Abstract] AND "gene*"[Title/Abstract]) AND ((fha[Filter]) AND (humans[Filter]))** Yielded: 151 Papers.

PPCM Search query: **(("Peripartum cardiomyopathy"[Title/Abstract] AND "gene*"[Title/Abstract]) AND ((fha[Filter]) AND (humans[Filter]))** Yielded: 121 Papers.

Search date: 28/11/2021.

We then performed the following steps: (1) analysis of the PMID and gene overlap of each dataset; (2) terms, authors and bio-entities were extracted from each dataset; (3) analysis of the overlap of genes in each dataset; (4) comparison of evolution patterns of each dataset; (5) development of network visualisation of co-authorship, co-terms, gene-disease co-occurrence and gene-chemical co-occurrence for PPCM; and (6) a focus on the PPCM data produced a map of gene importance vs specificity.

The use of AI and ML to identify key topics and themes in a bibliometric review is also a growing field in healthcare (Tran et al. 2019). Computer-aided techniques are not only limited to the natural language processing of scientific literature but are also being extended into detection (Adedinsewo et al. 2021; Awan et al. 2018), prediction (Angraal et al. 2020) and prognostication (McNamara et al. 2015) of the disease itself. The integration of these methodologies (Krittanawong et al. 2019) using data from the clinical records, genomics, transcriptomics and epigenetics will only improve over time.

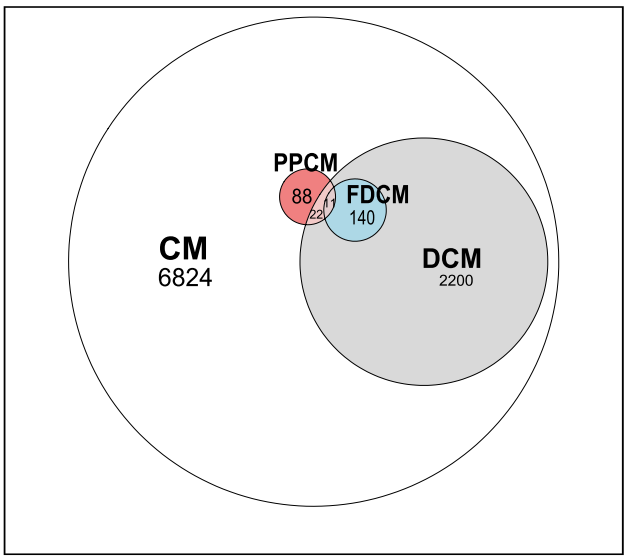

Fig. 6 Venn diagram showing the overlap of CM, DCM, FDCM, and PPCM journal articles by PMID in the literature

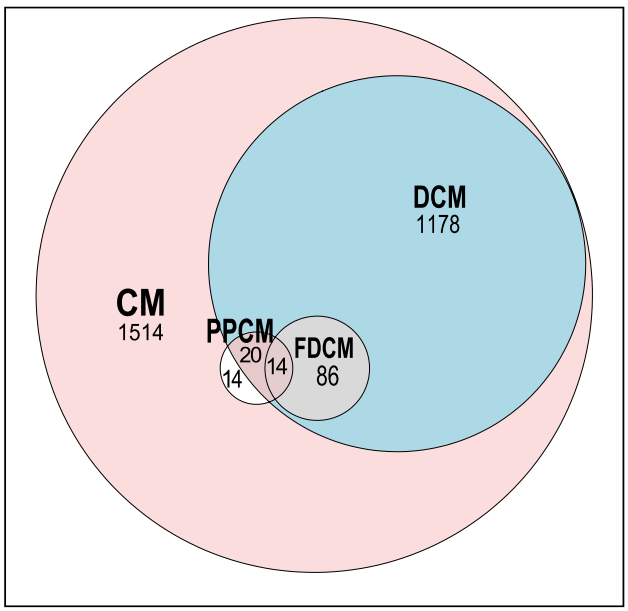

Fig. 7 Venn diagram showing the overlap of CM, DCM, FDCM and PPCM genes in the literature

Table 3 The unique fourteen PPCM genes not identified in DCM or FDCM research, their protein classes and location of expression

MappedID	Name	Protein class	Main location of expression
CFH	Complement factor H; CFH; ortholog		Liver and gallbladder, Muscle tissues, female tissues
CFI	Complement factor I; CFI; ortholog	Serine protease (PC00203)	Liver and gallbladder, Kidney and urinary bladder, female tissues
CTSD	Cathepsin D; CTSD; ortholog		Endocrine tissues, kidney and urinary bladder, Muscle tissues
GPT	Alanine aminotransferase 1; GPT; ortholog	Transaminase (PC00216)	Liver and gallbladder, Muscle tissues, female tissues
MMP23B	Matrix metalloproteinase-23; MMP23B; ortholog	Metalloprotease (PC00153)	Respiratory system, muscle tissues, female tissues
MR1	Major histocompatibility complex class I-related gene protein;MR1; ortholog	Major histocompatibility complex protein (PC00149)	Respiratory system, endocrine tissues, kidney and urinary bladder, bone marrow and lymphoid tissues
PTHLH	Parathyroid hormone-related protein; PTHLH; ortholog	Peptide hormone (PC00179)	Endocrine tissues, female tissues, proximal digestive tract
RLN1	Prorelaxin H1; RLN1; ortholog	Peptide hormone (PC00179)	No female tissues (male tissues)
RLN2	Prorelaxin H2; RLN2; ortholog	Peptide hormone (PC00179)	Female tissues, endocrine tissues, bone marrow and lymphoid tissues
SLC17A5	Sialin;SLC17A5; ortholog	Secondary carrier transporter (PC00258)	Endocrine tissues, kidney and urinary bladder, female tissue
SRI	Sorcin; SRI; ortholog	Calmodulin-related (PC00061)	Gastrointestinal tract, brain, female tissues
ST2,IL1RL1	Interleukin-1 receptor-like 1; IL1RL1; ortholog	Transmembrane signal receptor (PC00197)	Female tissue, respiratory system, kidney and urinary bladder
STAT5A	Signal transducer and activator of transcription 5A; STAT5A; ortholog	DNA-binding transcription factor (PC00218)	Connective and soft tissue, liver and gallbladder, bone marrow and lymphoid tissues
THBD	Thrombomodulin; THBD; ortholog	Transmembrane signal receptor (PC00197)	Respiratory system, skin, bone marrow and lymphoid tissues

Results

Figure 5 summarises the research output describing genomic-based investigations into the four types of cardiomyopathies that are exponentially increasing. It is clear that the total papers in genomics research on cardiomyopathy is greater than the aggregation of the DCM, FDCM and PPCM because other major classes of cardiomyopathies such as hypertrophic and restrictive CMs are also included. These areas suggest future topics for genomics comparative research. We also applied a year range filter between 2004 and 2021 to the search results to reflect recent developments in genomics. This time range was selected because genomics is a rapidly expanding field, particularly since the completion of the human genome sequencing project in 2001 (Venter et al. 2001).

PubMed ID analysis

Each article in PubMed carries a PubMed ID (PMID), which was used as a first step to compare the outputs of the four searches which were then analysed based on PMID overlap. This was performed via extraction and analysis in R. The Venn diagram (Fig. 6) highlights the overlap of papers based on the four search strategies. Several publications (e.g. Arany and Elkayam 2016; Ersbøll et al. 2016, Pearson et al. 2000) discuss how PPCM is closely correlated to both FDCM and DCM, but the data shown in this analysis (88 PPCM papers not linked to FDCM and DCM, compared with 11 linked to FDCM and DCM and 22 linked to DCM alone) suggest that there may be more differences than similarities between these classes of cardiomyopathy. This is contrary to the suggestion in recent papers covering these fields.

Gene analysis

Using natural language processing methodologies, we extracted the structured bio-entity data from the titles and abstracts of each query dataset. Further analysis in R was performed, and the Venn diagram (Fig. 7) highlights the genes and their overlap between CM, DCM, FDCM and PPCM in the literature. The genes listed for PPCM are as follows: TTN, HSP90B2P, PRL, CRP, CFH, CFI, THBD, CTSD, MMP23B, MMP1, MMP10, TIMP1, TIMP2,

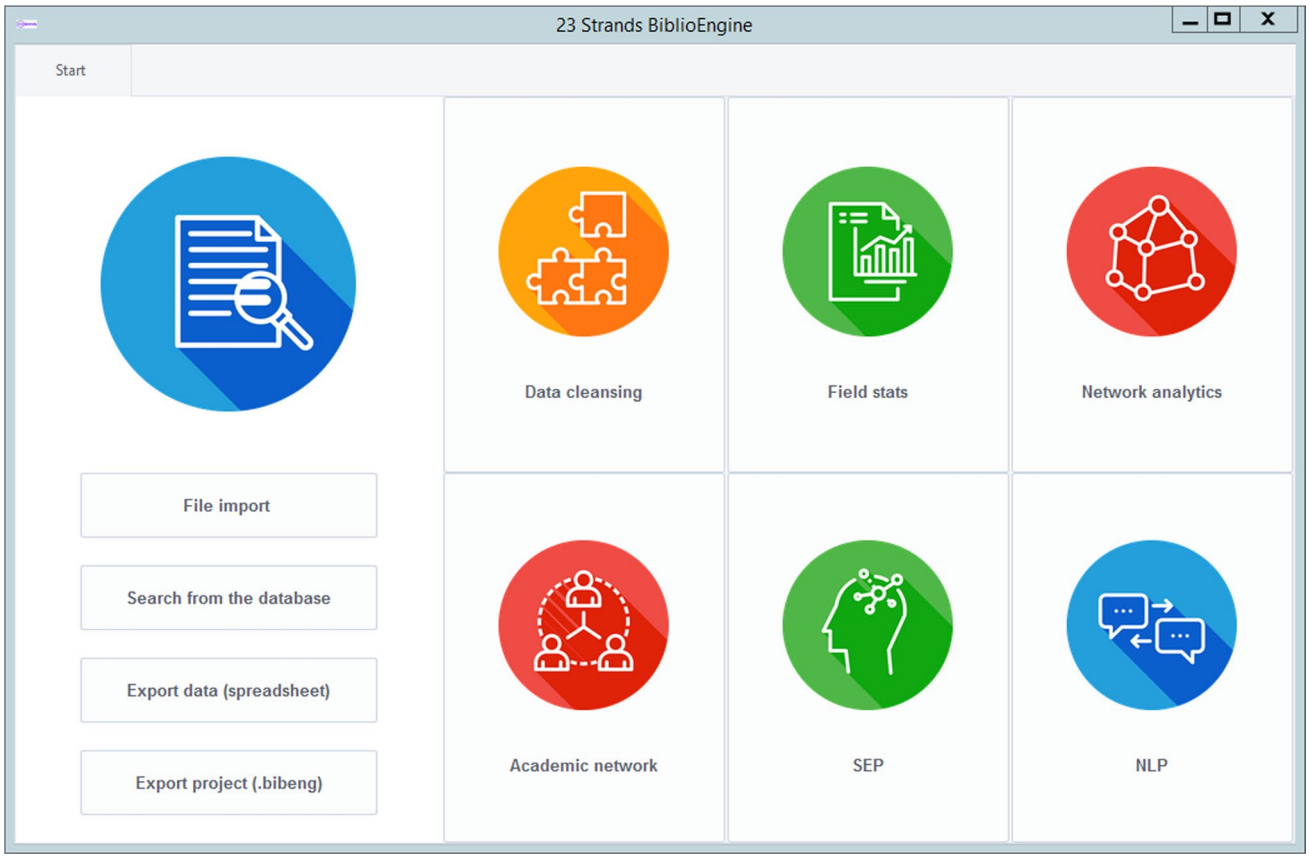

Fig. 8 Overview of the 23Strands' Biblioengine V2.2

TIMP4, ST2, LGALS3, SPP1, F2, KCNMB1, VEGFA, SLC17A5, GPT, ACE, RLN2, RLN1, STAT5A, STAT3, ADRB1, INS, EGFR, PLN, GNB3, EIF3K, BAG3, TNNC1, MYH7, IL6, IFNG, KCNH2, PTHLH, SCN5A, PSEN2, MYH6, TNNT2, MYBPC3, MR1, SRI and the miRNA gene, miR-146A. Of the above 48 genes (Fig. 7), 14 are not expressed in either FDCM or DCM and are therefore likely to be PPCM-specific: CFH, CFI, CTSD, GPT, MMP23B, MR1, PTHLH, RLN1, RLN2, SLC17A5, SRI, ST2(IL1RL1), STAT5A and THBD.

Once the gene data were extracted and identified from the three sets, a simple co-occurrence analysis was performed, and the genes peculiar to PPCM, but not DCM and FDCM, were highlighted. By using PANTHER (Mi et al. 2021) and the Human Protein Atlas (Thul and Lindskog 2018) expression services, RNAs expressed in tissues were identified in these 14 genes that were not expressed in DCM and FDCM (Table 3). They provide an opportunity to achieve novel insights into our understanding of the genes highlighted by Ware et al. (2016).

Contractile protein genes in the literature

The structural gene data extracted for each search strategy allowed a comparison with the proteins listed from Fig. 2. In addition to the list of major genes identified from the Human Protein Atlas, other family genes were identified from the literature itself. It was interesting to note that actin, myosin light chains, cardiac and troponin I did not appear to be researched in the PPCM but were researched in the other cardiomyopathies.

Using the Human Protein Atlas, the following list of genes were identified as are expressed proteins in cardiomyocytes: TTN, CTSD, LGALS3, KCNMB1, VEGFA, STAT3, ADRB1, EGFR, PLN, GNB3, EIF3K, BAG3, TNNC1, MYH7, KCNH2, PTHLH, SCN5A, MYH6, TNNT2, TPM1 and MYBPC3 (note: THBD is expressed in endothelial cells that are closely associated with cardiomyocytes). Additionally, the following were detected in healthy human LV tissue as RNA transcripts but are not detected at the protein level: CFH, MMP23B, SRI, F2, STAT5A, STAT3, TIMP1, TIMP2 and TIMP4.

The following genes were not detected in healthy human LV tissue at either the protein or RNA level: PRL, CRP, SLC17A5, CFI, MMP1, MMP10, GPT, IFNG and HSP90B2P. An example of the importance of identifying novel genes in the research is the PTHLH gene (Horne et al. 2011) that was shown by genome-wide association (GWAS) to provide a potential genetic basis for PPCM.

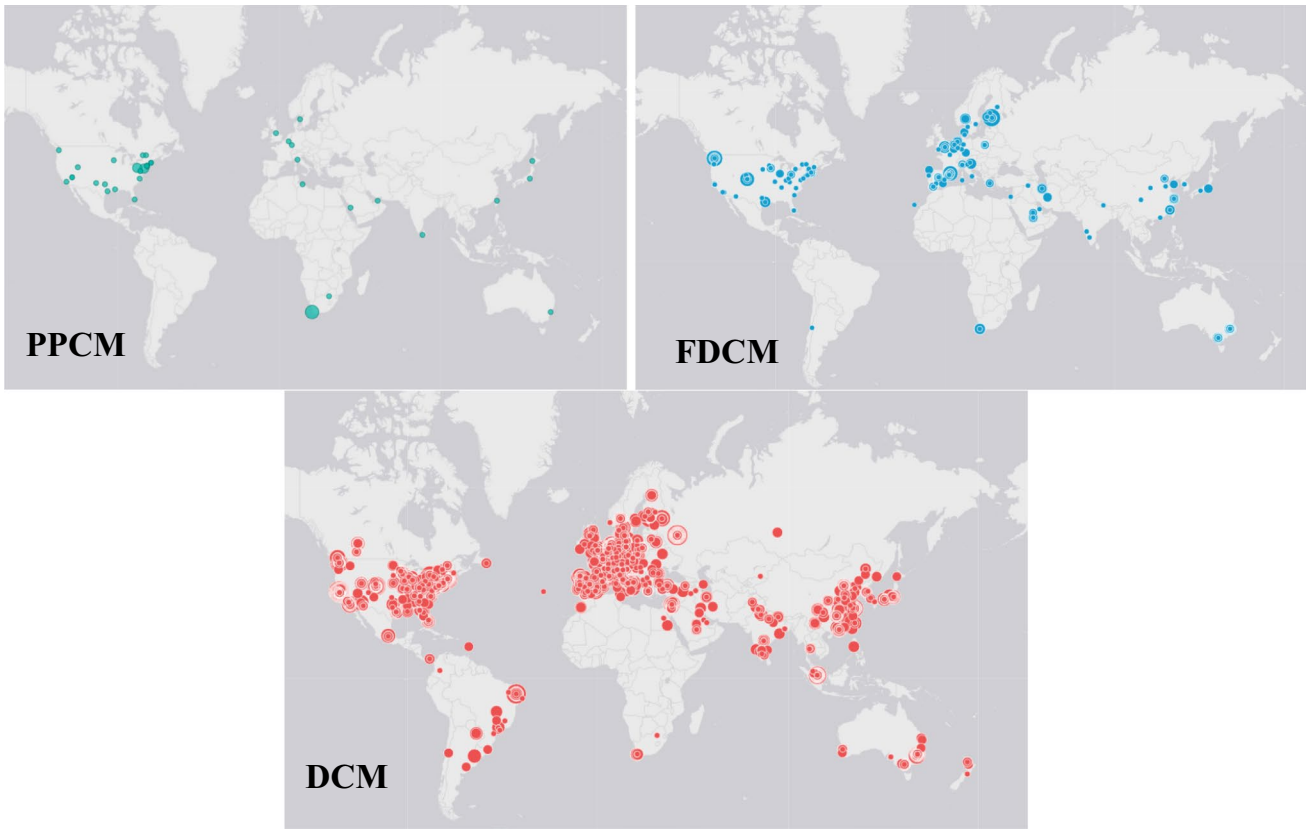

Fig. 9 The global locations of primary author affiliations in genomics research shown across PPCM, FDCM and DCM, the size of bubbles represent counts of publications

Once the analysis of overlap was completed between the different cardiomyopathies, a more detailed analysis of PPCM literature was performed.

Peripartum cardiomyopathy-specific insights

The set of 121 PPCM related genomics papers was then analysed using techniques developed by the team and described in detail in Zhang et al. (2017, 2020) and Wu et al. (2021). This analysis was completed using the 23Strands' BiblioEngine (Fig. 8), a set of natural language processing AI and ML empowered tools designed specifically for the extraction, cleaning and analysis of scientific journal articles.

Research locations

The first descriptive data extracted from the set of journal articles was the cleaned primary affiliations of the authors, which were then geo-located using ESRI services (Fig. 9). The map shows that PPCM genomics research is spread across most of Asia, Africa and Europe, but more research is needed if we are to understand differences in the incidence of PPCM, particularly in women of African descent (Irizarry et al. 2017).

Scientific evolution pathways

Once terms were extracted from the papers and cleaned, a Scientific Evolutionary Pathway (Zhang et al. 2017) was generated. Even though the number of papers was limited to 121 articles, the technique was useful for understanding the development of concepts around PPCM and molecular and genomic links over time.

The scientific evolutionary pathway of MeSH terms in PPCM (Fig. 10) shows how the initial topic of "pregnancy complication, cardiovascular" (van Mook and Peeters 2005) evolved systematically into specific insights into "hyperprolactinemia" (Shelly et al. 2012), "puerperal disorders" (e.g. Morales et al. 2010), "cardiac myocytes" (Harakalova et al. 2015), "registries" (Haghikia et al. 2013), "mutation" (Fish et al. 2016) and "hypertension, pregnancy-induced" (Katsuragi et al. 2019).

The topic of pregnancy-induced hypertension (otherwise known as pre-eclampsia) was particularly of interest because

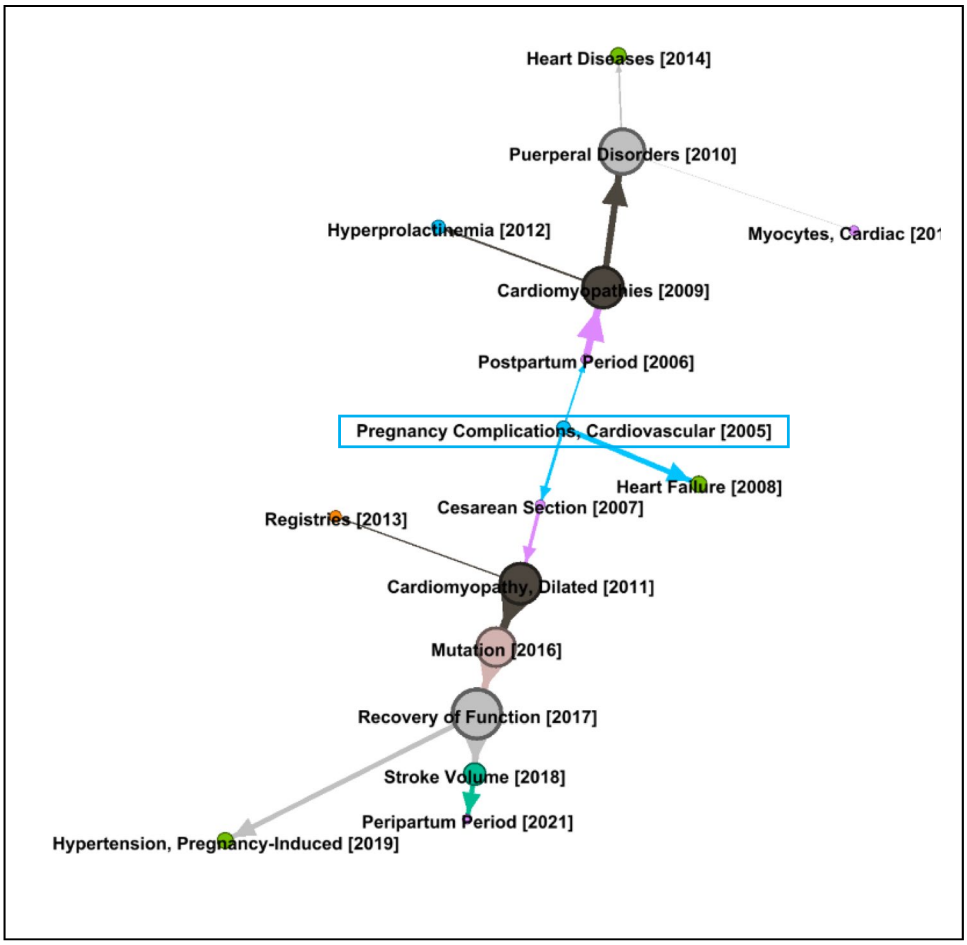

Fig. 10 Scientific Evolutionary Pathway of PPCM and genomics, note how the topic evolution extends up and down from the first topic of pregnancy complications (Cardiovascular (2005))

the definition of DCM clearly states that the heart cannot be under undue stress or load for DCM to be diagnosed, indicating a possible point of difference between PPCM and DCM classification.

Co-occurrence analysis and visualisations

Following the geolocation mapping of the cleaned affiliations, further co-occurrence analysis was performed on authors collaborating on papers, affiliation collaborations (authors institutions), gene plus sub-disease links and gene plus chemical co-occurrence.

It is clear from Figs. 11 and 12 that the research on PPCM is led by key individuals and their affiliations. In particular, Hilfiker-Kleiner (Sliwa et al. 2021) and Sliwa (Spracklen et al. 2021a, b) have made significant contributions to the genomics of PPCM and had a global impact and collaboration across Africa, Europe and the Americas. A smaller cluster of research occurred in Japan led by Ikeda (Kamiya et al. 2016) and his team.

The co-occurrence analysis (Fig. 13) shows the link between genes and sub-categories of diseases resulting in a significant number of clinical entities and/or comorbidities connected with PPCM. Apart from the expected links of heart disease, heart failure and death there are interesting comorbidities including arrhythmias, pre-eclampsia, respiratory disease, inflammation, autoimmunity, anaemia, lupus and pregnancy-related symptoms such as lactation, gestational diabetes and orthopnoea.

Additionally, the gene plus chemical co-occurrence (Fig. 14) shows two main clusters relating to treatments for hyperprolactinemia, decreasing the amount of prolactin in the body and possible links with inflammation compared with the second cluster relating to cardiogenic shock treatments, such as dobutamine.

Figure 15 illustrates early analysis on the extracted genes from the PPCM literature, a process described in detail in Wu et al. (2021). This graph allows us to differentiate genes based on an analysis of the geometries in co-occurrence in the literature. From this analysis, we identified two main clusters; the group highlighted in the red box indicates genes that have a range of specificity with PPCM and other diseases but are central to the PPCM topic. The genes outside the red box also have a range of specificity to the PPCM topic but are less central to the topic. This method of multivariate co-occurrence analysis allows us to use a different

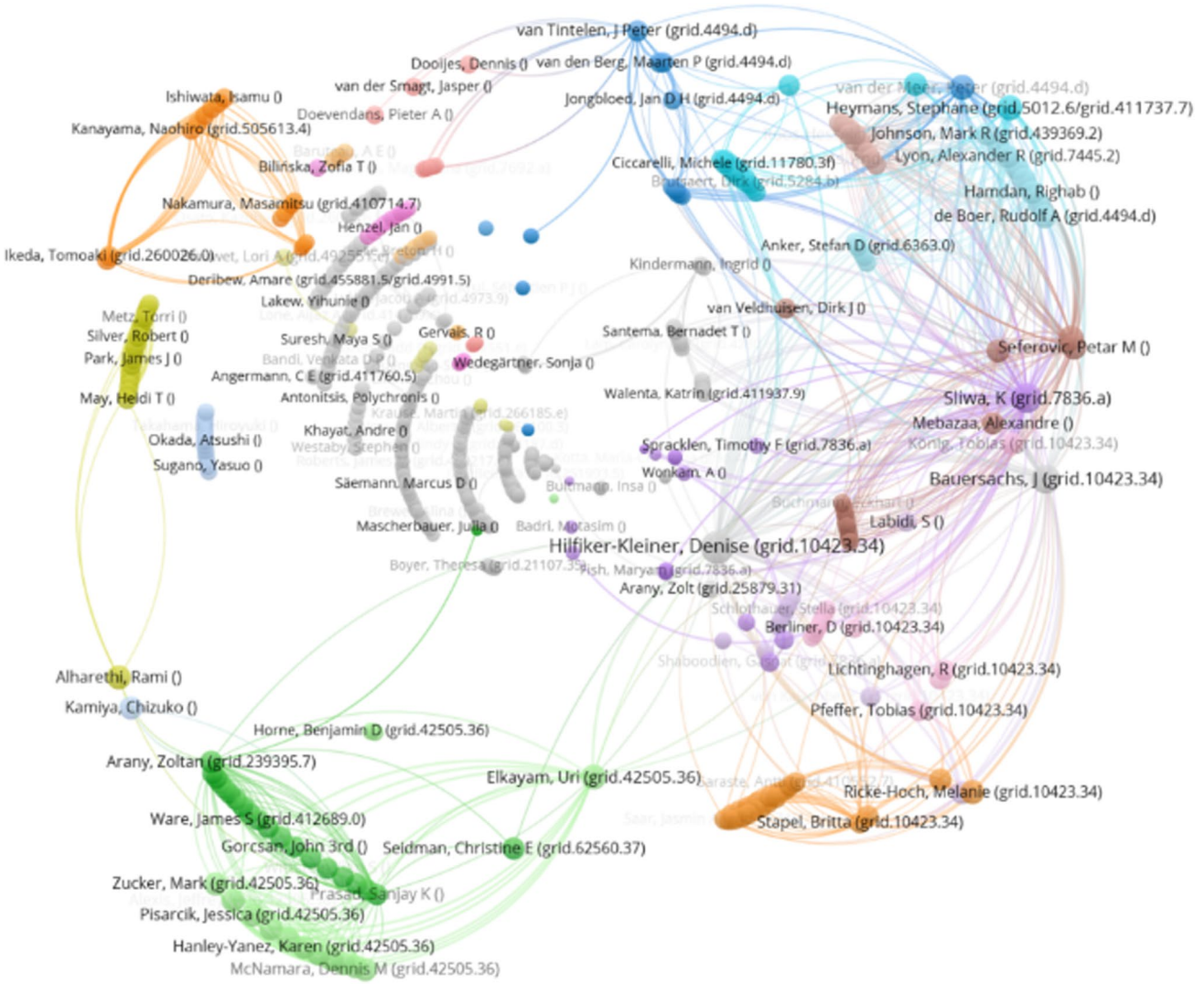

Fig. 11 A study of author co-occurrence revealed clusters around Sliwa (purple), Hilfiker-Kleiner (grey), Elkayam (green), Ricke-Hoch (orange) and Ikeda in the top left (light orange). The size of the bubbles matches the number of articles published. The grid numbers have been used to disambiguate the authors

lens to classify genes extracted from the PPCM literature. Through this method, we can see a marked difference in genes that are unique to PPCM (Table 1), and genes highlighted with the additional centrality analysis.

An example of a gene identified as unique to PPCM research publications (compared with DCM and FDCM) is CFI (complement factor I highlighted by a green box in Fig. 15) that encodes a serine protease that is essential for regulating the complement cascade. CFI is referred to in a single paper, Amari Chinchilla et al. (2020), and was identified in this GWAS study as having a possible association with PPCM. In the paper, it was also linked to other diseases such as pregnancy-associated haemolytic uremic syndrome (P-aHUS), thrombotic thrombocytopenic purpura, preeclampsia/haemolysis, elevated liver enzymes, low platelets (HELLP) and systemic lupus erythematosus/antiphospholipid syndrome. The analysis highlights how CFI differs significantly from the prolactin gene (PRL also highlighted by a green box in Fig. 15). The most outstanding gene from the centrality analysis is described in eight papers, including Ersbøll et al. (2016), which all focus specifically on how PRL impacts PPCM.

In Table 4, we provide a total list of the 48 genes shown in Fig. 15 and again used the Human Protein Atlas (Thul and Lindskog 2018) to describe the protein classes and tissue RNA expression values in heart muscle (Heart RNA), as well as specific expression in cardiomyocytes (CMy RNA).

In summary, this study has identified genes for further investigation. For example, the vasculo-hormonal pathogenesis hypothesis of PPCM (Fig. 16) proposed to explain

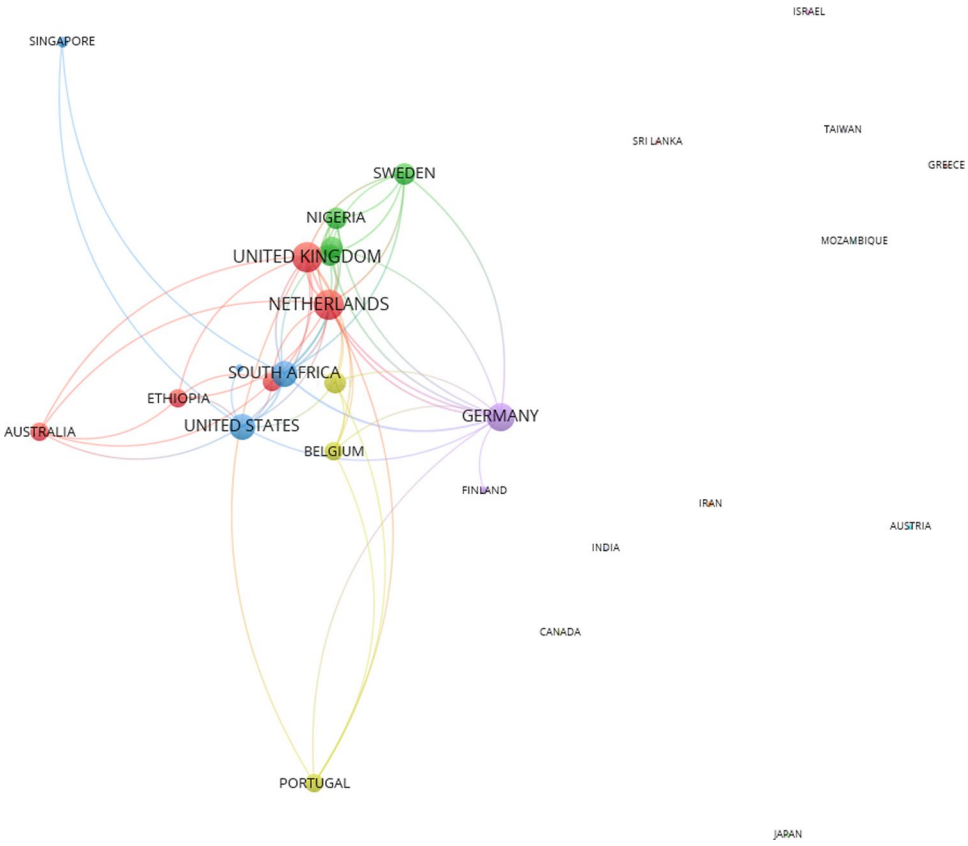

Fig. 12 Country co-occurrence; this clearly shows collaborative research across Europe, Africa and the United States and a number of islands of research including Japan

the pathophysiology of PPCM (Adapted from Bello and Arany 2015), namely, that during the peripartum period, there is an increased secretion of hormones from the pituitary (e.g. prolactin) and from the placenta (sFLT1). These changes are believed to contribute to the underlying cardiac dysfunction in PPCM.

Discussion, conclusion and future work

By performing an AI, including ML, empowered semi-automated analysis of the scientific literature, it became clear that a large amount of data could be processed quickly that produced new insights into this rare disease and the connection between PPCM and genomics. This would be hugely time-consuming using manual systematic reviews. It would be almost impossible to do a similar quantitative, comparative analysis of the gene research across the set of 9,285 CM papers and genomics-related journal articles within a useful timeframe. Using only publicly available data extracted from PubMed, we have shown how it is possible to answer questions as shown below.

What specific genes should be considered, based on current knowledge for PPCM?

PPCM is a subset of cardiomyopathy albeit with a low number of genomics research articles in the US NLM literature ($n = 121$). It is clear that we are only in the early stages of understanding the disease-gene relationships. Fourteen genes closely associated in the current literature with PPCM have been identified by the centrality analysis (PRL, RLN2, RLN1, PLN, ST2, CTSD, F2, ACE, STAT3, TTN, SPP1, LGALS3, GNB3, SRI) and help to form the basis of future studies regarding how they may or may not play a part in the development of PPCM. Such work could also help improve the diagnosis and management of this disease.

Using this big data approach, it is not only possible to focus on the signals that provide an alternative method for identifying the number of key genes in the research, but also it is possible to look at the full dataset of current CM-related genes. These computer-aided AI techniques allowed us to quickly explore interrelationships such as potential chemical interactions, as well as the transcriptome and proteome in the context of PPCM and other cardiomyopathies.

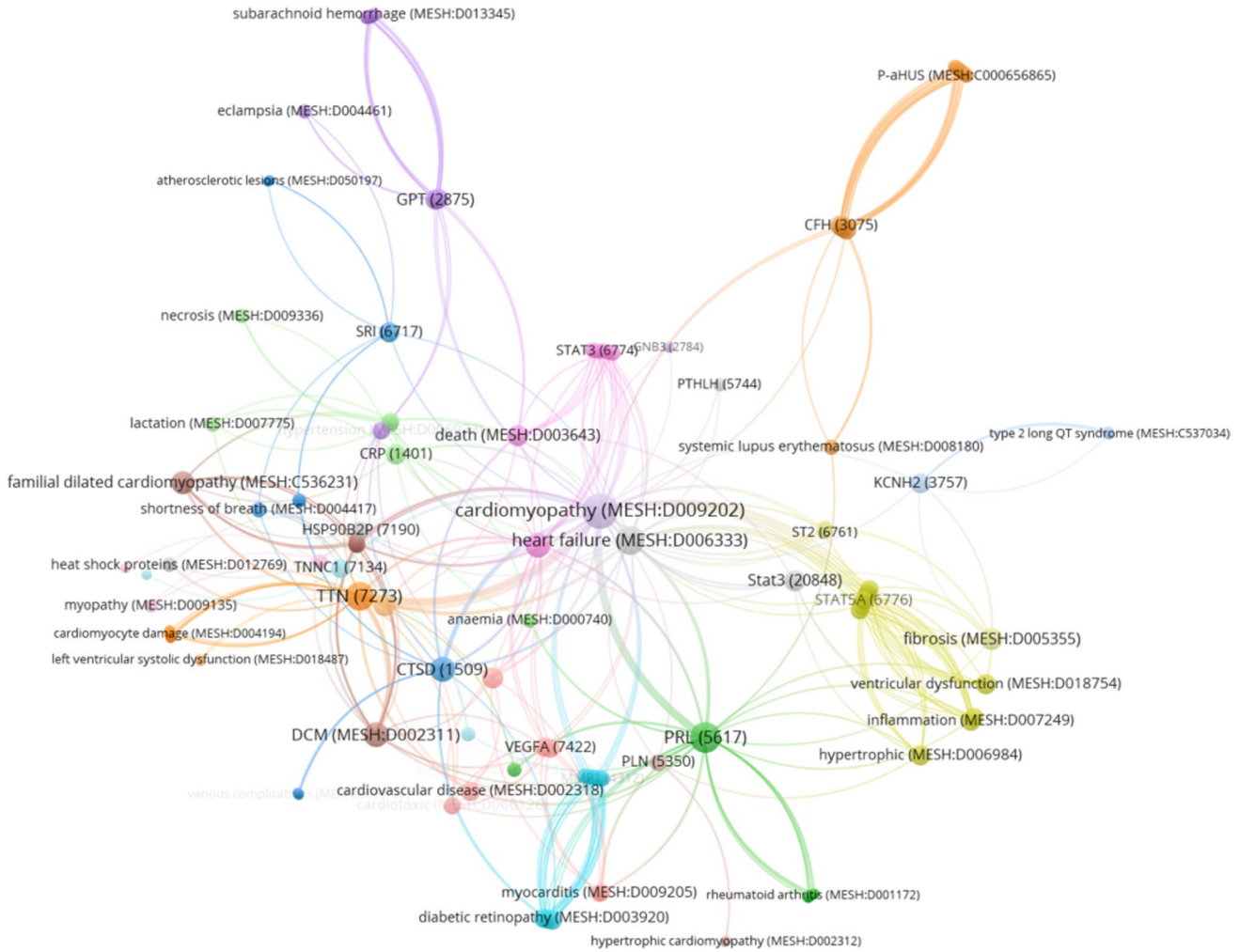

Fig. 13 Gene plus sub-disease co-occurrence shows the multitude of co-morbidities identified in the research title and abstracts and the co-occurrence with the major genes

Where are the major affiliations and who are the authors of PPCM research?

We clearly identified where the main research on PPCM is currently being performed and identified future collaborative research possibilities across India, China and other parts of Asia Pacific. This exercise may help to develop a more equitable, global research effort on PPCM.

How can AI help in reviewing other cardiomyopathies and aid in understanding PPCM?

This research focused on the novel use of AI and ML technologies to produce a review comparing the several cardiomyopathies. The methods used in developing this research involve data extraction, Natural Language Processing, disambiguation, network and graphical analysis that reveal the utility of these methods in investigating multidimensional diseases. We believe there are significant opportunities for the future application of other AI empowered techniques and algorithms, such as other natural language processing of genomics-related deep neural networks as well as integrating other dimensions such as medical records and data analysis of patients' deep phenotype. Each of these methods has the potential to make a significant impact on discovering new pathways for researchers and decision support mechanisms for the clinicians managing these patients.

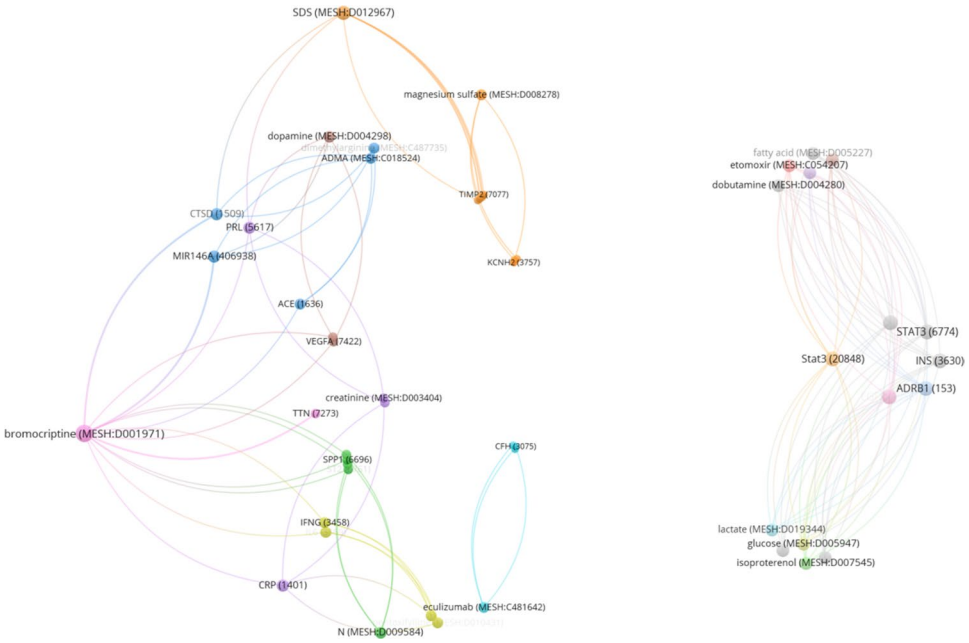

Fig. 14 Gene plus chemical co-occurrence highlighting two clusters of inflammation and cardiogenic shock

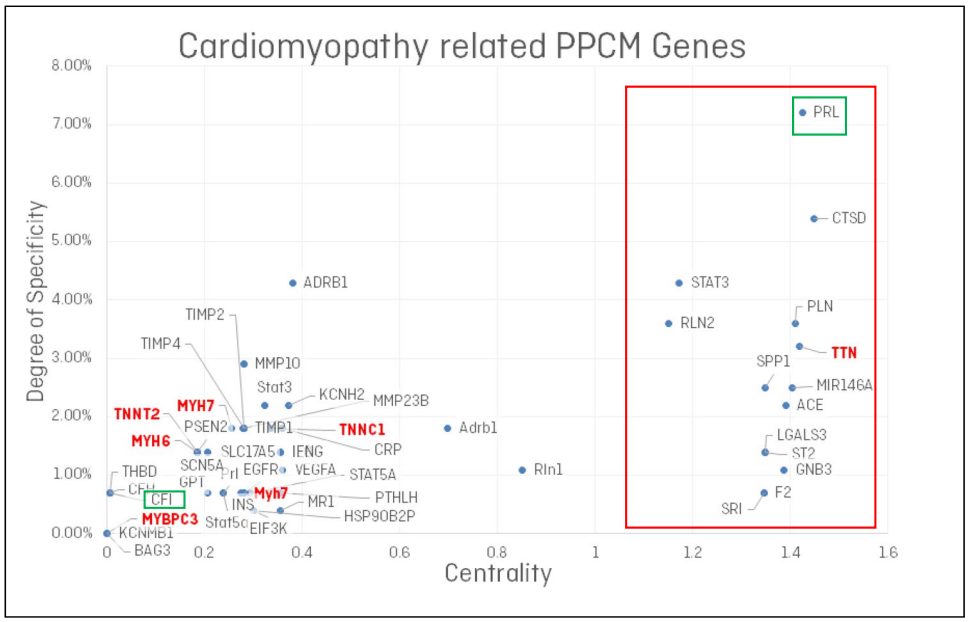

Fig. 15 Centrality analysis of PPCM Genes with relationship to cardiomyopathy, cardiac contractile proteins highlighted in red. Centrality is measured as a composite of the influence (degree centrality), topological distance (closeness centrality) and the shortest path (betweenness centrality) of the gene nodes relating to the topic of "cardiomyopathy" in the scientific literature on PPCM (Freeman et al. 1979)

What can we do to understand more clearly the challenges that arise in reviewing a limited researched space?

During this process, we identified a number of further research areas that could be undertaken in this space of genomics plus disease relationship analysis, and additional work is planned in linking other third-party datasets and simulation and modelling techniques.

The future of AI in disease-genomics links

There remains significant potential to harness these computer-aided techniques for scientific knowledge. However, before algorithms such as transfer learning can be used to gain deeper understanding of the relationships between genomics and rare diseases, a first necessary step is to understand the current published literature. The comparison of PPCM

Table 4 Genes identified in the PPCM literature, their protein classes and main location of expression. Bold indicates centrality-based importance. Red indicates related contractile proteins

Symbol	Name	Heart RNA	CMy RNA	Protein class	Biological process
PRL	**Prolactin**	**0**	**0**	**Cancer-related genes, plasma proteins, predicted secreted proteins**	**Lactation**
TTN	**Titin**	**136.5**	**19,161.7**	**Disease-related genes, enzymes, plasma proteins, potential drug targets, predicted intracellular proteins, predicted membrane proteins**	
ACE	Angiotensin I converting enzyme	4.7	1.4	Candidate cardiovascular disease genes, CD markers, disease-related genes, enzymes, FDA-approved drug targets, plasma proteins, predicted intracellular proteins, predicted membrane proteins, predicted-secreted proteins	
CTSD	**Cathepsin D**	**86**	**226.4**	**Cancer-related genes, Disease related genes, Enzymes, Plasma proteins, Potential drug targets, Predicted intracellular proteins, Predicted secreted proteins**	
GNB3	**G protein subunit beta 3**	**5.4**	**4.6**	**Disease-related genes, plasma proteins, predicted intracellular proteins, RAS pathway-related proteins**	
PLN	**Phospholamban**	**337.5**	**2230.5**	**Disease-related genes, potential drug targets, predicted membrane proteins, transporters**	
RLN2	**Relaxin 2**	**0.6**	**1.6**	**Predicted intracellular proteins, predicted secreted proteins**	
STAT3	**Signal transducer and activator of transcription 3**	**82.9**	**46.2**	**Cancer-related genes, disease-related genes, plasma proteins, predicted intracellular proteins, transcription factors**	**Host-virus interaction, Transcription, Transcription regulation**
ADRB1	Adrenoceptor beta 1	48	80	FDA-approved drug targets, G-protein coupled receptors, predicted membrane proteins, transporters	
SPP1	**Secreted phosphoprotein 1**	**2.6**	**4.3**	**Cancer-related genes, plasma proteins, predicted secreted proteins**	**Biomineralization, cell adhesion**
MMP1	Matrix metallopeptidase 1	0.4	0	Cancer-related genes, candidate cardiovascular disease genes, enzymes, FDA-approved drug targets, plasma proteins, predicted secreted proteins	Collagen degradation, host-virus interaction
MMP10	Matrix metallopeptidase 10	0.1	0	Cancer-related genes, enzymes, FDA-approved drug targets, predicted secreted proteins	Collagen degradation
MMP23B	Matrix metallopeptidase 23B	14.4	9	Enzymes, FDA-approved drug targets, predicted intracellular proteins	
MYH7	Myosin heavy chain 7	644.7	8327.1	Candidate cardiovascular disease genes, disease-related genes, plasma proteins, predicted intracellular proteins	
TNNC1	Troponin C1, slow skeletal and cardiac type	431.6	2506.5	Disease-related genes, FDA-approved drug targets, plasma proteins, predicted intracellular proteins	
CRP	C-reactive protein	0.1	0	Cancer-related genes, candidate cardiovascular disease genes, plasma proteins, predicted intracellular proteins, predicted secreted proteins	Acute phase
ST2; IL1RL1	**Interleukin 1 receptor like 1**	**16.6**	**1.8**	**Candidate cardiovascular disease genes, plasma proteins, predicted intracellular proteins, predicted membrane proteins, predicted secreted proteins**	

Table 4 (continued)

Symbol	Name	Heart RNA	CMy RNA	Protein class	Biological process
KCNH2	Potassium voltage-gated channel subfamily H member 2	24.9	76.1	Disease-related genes, FDA-approved drug targets, plasma proteins, predicted membrane proteins, transporters, voltage-gated ion channels	Ion transport, potassium transport, transport
LGALS3	**Galectin 3**	**42.3**	**183.5**	**Cancer-related genes, plasma proteins, predicted intracellular proteins, predicted-secreted proteins**	**Differentiation, immunity, innate immunity, mRNA processing, mRNA splicing**
PTHLH	Parathyroid hormone like hormone	4	0.3	Cancer-related genes, disease-related genes, predicted secreted proteins	
SLC17A5	Solute carrier family 17 member 5	6.2	5.3	Disease-related genes, predicted membrane proteins	Amino-acid transport, symport, transport
STAT5A	Signal transducer and activator of transcription 5A	9.8	20.2	Cancer-related genes, plasma proteins, predicted intracellular proteins, transcription factors	Lactation, transcription, transcription regulation
BAG3	BAG cochaperone 3	38.2	186.4	Disease-related genes, plasma proteins, predicted intracellular proteins	Apoptosis
CFH	Complement factor H	47.3	6.2	Cancer-related genes, disease-related genes, plasma proteins, predicted secreted proteins	Complement alternate pathway, host-virus interaction, immunity, Innate immunity
CFI	Complement factor I	6.2	1.9	Disease-related genes, enzymes, plasma proteins, potential drug targets, predicted secreted proteins	Complement pathway, host-virus interaction, Immunity, innate immunity
EGFR	Epidermal growth factor receptor	4.8	7.3	Cancer-related genes, disease-related genes, enzymes, FDA-approved drug targets, plasma proteins, predicted intracellular proteins, predicted membrane proteins, predicted secreted proteins, RAS pathway-related proteins	Host-virus interaction
EIF3K	Eukaryotic translation initiation factor 3 subunit K	84	408.5	Predicted intracellular proteins	Protein biosynthesis
F2	**Coagulation factor II, thrombin**	**0**	**0.6**	**Cancer-related genes, candidate cardiovascular disease genes, disease-related genes, enzymes, FDA-approved drug targets, plasma proteins, predicted intracellular proteins, predicted secreted proteins**	**Acute phase, blood coagulation, haemostasis**
GPT	Glutamic-pyruvic transaminase	16.9	6.5	Enzymes, plasma proteins, predicted intracellular proteins	
IFNG	Interferon gamma	1	0.2	Cancer-related genes, disease-related genes, FDA-approved drug targets, plasma proteins, predicted secreted proteins	Antiviral defense, growth regulation
IL6	Interleukin 6	21.6	0.9	Cancer-related genes, candidate cardiovascular disease genes, disease-related genes, FDA-approved drug targets, plasma proteins, predicted intracellular proteins, predicted secreted proteins	Acute phase
INS	Insulin	0.1	0	Cancer-related genes, candidate cardiovascular disease genes, disease-related genes, plasma proteins, predicted secreted proteins, RAS pathway-related proteins	Carbohydrate metabolism, glucose metabolism
KCNMB1	Potassium calcium-activated channel subfamily M regulatory beta subunit 1	2.6	2.6	FDA-approved drug targets, predicted membrane proteins, transporters	Ion transport, transport
MR1	Major histocompatibility complex, class I-related	8.9	3.9	FDA-approved drug targets, predicted intracellular proteins, predicted membrane proteins	Immunity, innate immunity
MYBPC3	Myosin binding protein C3	199.6	1683.6	Disease-related genes, plasma proteins, predicted intracellular proteins	Cell adhesion
MYH6	Myosin heavy chain 6	313.7	5369.5	Candidate cardiovascular disease genes, disease-related genes, plasma proteins, predicted intracellular proteins	

Table 4 (continued)

Symbol	Name	Heart RNA	CMy RNA	Protein class	Biological process
PSEN2	Presenilin 2	9.7	7.5	Disease-related genes, enzymes, potential drug targets, predicted intracellular proteins, predicted membrane proteins, transporters	Notch signalling pathway
RLN1	Relaxin 1	0	0.1	Plasma proteins, predicted secreted proteins	
SCN5A	Sodium voltage-gated channel alpha subunit 5	67.3	66.1	Disease-related genes, FDA-approved drug targets, plasma proteins, predicted intracellular proteins, predicted membrane proteins, transporters, voltage-gated ion channels	Ion transport, sodium transport, transport
SRI	**Sorcin**	**19.4**	**95.1**	**Predicted intracellular proteins**	
THBD	Thrombomodulin	24.3	9	Candidate cardiovascular disease genes, CD markers, disease-related genes, predicted membrane proteins	Blood coagulation, haemostasis
TIMP1	TIMP metallopeptidase inhibitor 1	105.7	8.1	Cancer-related genes, candidate cardiovascular disease genes, plasma proteins, predicted intracellular proteins, predicted secreted proteins	
TIMP2	TIMP metallopeptidase inhibitor 2	38.5	35.8	Cancer-related genes, candidate cardiovascular disease genes, plasma proteins, predicted intracellular proteins, predicted secreted proteins	
TIMP4	TIMP metallopeptidase inhibitor 4	10.9	4.8	Plasma proteins, predicted secreted proteins	
TNNT2	Troponin T2, cardiac type	679.2	11,884.8	Candidate cardiovascular disease genes, disease-related genes, predicted intracellular proteins	
VEGFA	Vascular endothelial growth factor A	57	146.9	Cancer-related genes, candidate cardiovascular disease genes, disease related genes, FDA-approved drug targets, plasma proteins, predicted intracellular proteins, predicted secreted proteins, RAS pathway-related proteins	Angiogenesis, differentiation

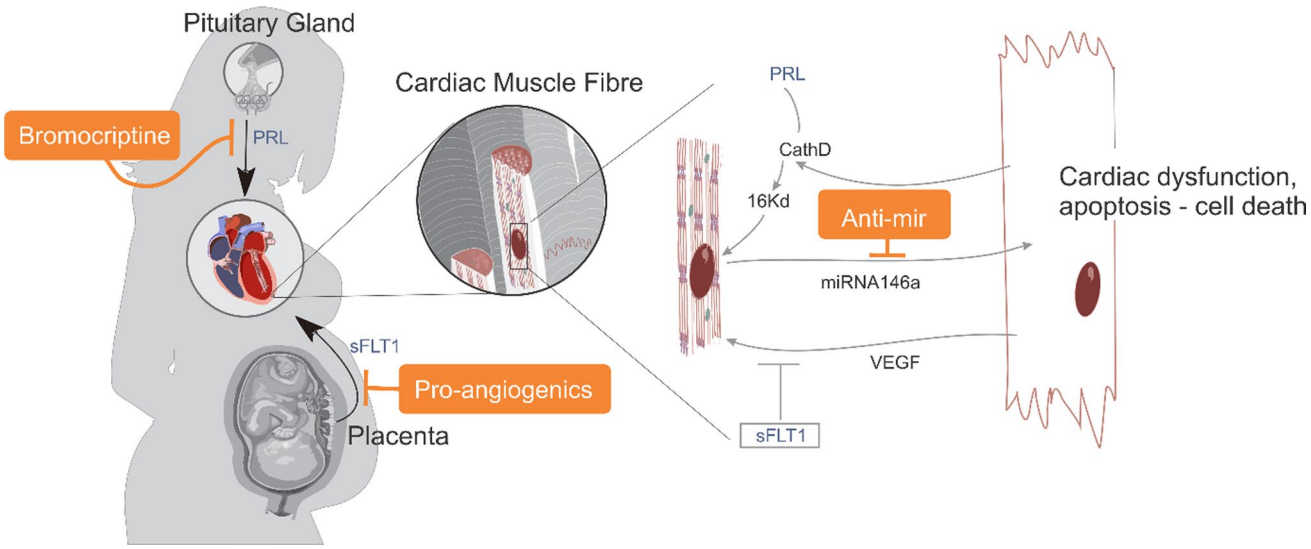

Fig. 16 A vasculo-hormonal hypothesis of the pathophysiology of PPCM (adapted from Bello and Arany 2015), during the peripartum period there is an increased secretion of hormones from the pituitary (e.g. prolactin (PRL)) and placenta (e.g. sFLT1). This change is believed to contribute to the underlying cardiac dysfunction in PPCM

genomics with other cardiomyopathies offers a good example of this, and it was the aim of this paper to show how this first step of AI empowered review of the current scientific landscape could be achieved.

Conclusions relating to PPCM

It is clear that PPCM is a disease that effects a complex system of both the heart, hormone and haemodynamic vascular systems of genes, transcriptome and epigenetics. Using the AI-empowered bibliometrics approach, we extracted the titles and abstracts to identify the main genes and their relating proteins, involved in regulating dysfunction. In addition, this analysis automatically linked genes extracted with the developing hypothesis of hormonal and metabolic stress related causes of PPCM.

A discussion on the strengths and weaknesses of this review technique

This study is a prototype of Bibliometrics, empowered with AI algorithms to identify signals used to profile the research landscape on a disease such as PPCM.

This research highlighted both a number of strengths and weaknesses of this approach:

Strengths

1. A significant number of publications can be processed in a relatively short amount of time.
2. Extraction of structural features of these topics can be automated.
3. Related disease features can be compared with ease.
4. Linkages created though common authors, affiliations, publication and extracted features can be used to visualise and analyse data.
5. This “big data” approach can provide a foundation for future algorithms for predicting future research and collaboration.

Weaknesses

6. The outputs of the analysis are highly dependent on the search strategy, and we believe that it is important to review the outputs of bibliometrics using manual expert validation.
7. Signal-to-noise is hard to determine and quantify, but will be critical to identify signals versus noise and other weaknesses in the algorithms.
8. There is a need to understand the correlation versus causation which is difficult to establish using Abstract alone.

9. Access to the full text of journals is limited due to a large amount of research not accessible due to commercial access constraints.

Author contribution MG is the founder and CEO of 23 Strands Pty Ltd (Australia) and was the primary author of this manuscript. HL is the Chief Data Officer of 23Strands and co-ordinated the development and implementation of the techniques shown with MW, YJ and JL from the Australian Artificial Intelligence Institute at the University of Technology, Sydney. DV is the Chief Medical Advisor, and ST is the Chief Strategy Officer at 23Strands and both were involved in writing the manuscript. CdR is from the Victor Chang Cardiac Research Institute and conceived of the work and was involved in writing the manuscript.

Declarations

Consent to participate Not applicable.

Consent for publication The authors give consent to publish this manuscript.

Conflict of interest MG, HL, DV, and ST are employees of 23 Strands Pty Ltd (Australia), a privately held company, but their employment does not alter the authors' adherence to the publication policies.

References

Adedinsewo DA, Johnson PW, Douglass EJ, Attia IZ, Phillips SD, Goswami RM, Yamani MH, Connolly HM, Rose CH, Sharpe EE, Blauwet L, Lopez-Jimenez F, Friedman PA, Carter RE, Noseworthy PA (2021) Detecting cardiomyopathies in pregnancy and the postpartum period with an electrocardiogram-based deep learning model. Eur Heart J Digit Health 2:586–596. https://doi.org/10.1093/ehjdh/ztab078

Amari Chinchilla K, Vijayan M, Taveras Garcia B, Jim B (2020) Complement-mediated disorders in pregnancy. Adv Chronic Kidney Dis 27:155–164. https://doi.org/10.1053/j.ackd.2020.01.002

Angraal S, Mortazavi BJ, Gupta A, Khera R, Ahmad T, Desai NR, Jacoby DL, Masoudi FA, Spertus JA, Krumholz HM (2020) Machine learning prediction of mortality and hospitalization in heart failure with reserved ejection fraction. JACC Heart Fail 8:12–21. https://doi.org/10.1016/j.jchf.2019.06.013

Arany Z, Elkayam U (2016) Peripartum cardiomyopathy. Circulation 133:1397–1409. https://doi.org/10.1161/CIRCULATIONAHA.115.020491

Awan SE, Sohel F, Sanfilippo FM, Bennamoun M, Dwivedi G (2018) Machine learning in heart failure: ready for prime time. Curr Opin Cardiol 33:190–195. https://doi.org/10.1097/HCO.0000000000000491

Bello NA, Arany Z (2015) Molecular mechanisms of peripartum cardiomyopathy: a vascular/hormonal hypothesis. Trends Cardiovasc Med 25:499–504. https://doi.org/10.1016/j.tcm.2015.01.004

Bollen IAE, Van Deel ED, Kuster DWD, Van Der Velden J (2015) Peripartum cardiomyopathy and dilated cardiomyopathy: different at heart. Front Physiol 5.https://doi.org/10.3389/fphys.2014.00531

Charron P, Elliott PM, Gimeno JR, Caforio ALP, Kaski JP, Tavazzi L, Tendera M, Maupain C, Laroche C, Rubis P, Jurcut R, Calò L, Heliö TM, Sinagra G, Zdravkovic M, Kavoliūnienė A, Felix SB, Grzybowski J, Losi M-A, Asselbergs FW, García-Pinilla JM, Salazar-Mendiguchia J, Mizia-Stec K, Maggioni AP, Cardiomyopathy Registry Investigators EORP, Anastasakis A, Biagini E, Bilinska Z, Castro FJ, Celutkiene J, Chakova N, Chmielewski P, Drago F, Frigy A, Frustaci A, Garcia-Pavia P, Hinic S, Kindermann I, Limongelli G, Medrano C, Monserrat L, Olusegun-Joseph A, Ripoll-Vera T, Rocha Lopes L, Saad A, Sala S, Seferovic PM, Sepp R, Urbano-Moral JA, Villacorta E, Wybraniec M, Yotti R, Zachara E, Zorio E (2018) The cardiomyopathy registry of the EURObservational Research Programme of the European Society of Cardiology: baseline data and contemporary management of adult patients with cardiomyopathies. Eur Heart J 39:1784–1793. https://doi.org/10.1093/eurheartj/ehx819

Davis MB, Arany Z, McNamara DM, Goland S, Elkayam U (2020) Peripartum cardiomyopathy. J Am Coll Cardiol 75:207–221. https://doi.org/10.1016/j.jacc.2019.11.014

Elliott P, Andersson B, Arbustini E, Bilinska Z, Cecchi F, Charron P, Dubourg O, Kuhl U, Maisch B, McKenna WJ, Monserrat L, Pankuweit S, Rapezzi C, Seferovic P, Tavazzi L, Keren A (2007) Classification of the cardiomyopathies: a position statement from the European society of cardiology working group on myocardial and pericardial diseases. Eur Heart J 29:270–276. https://doi.org/10.1093/eurheartj/ehm342

Ersbøll AS, Damm P, Gustafsson F, Vejlstrup NG, Johansen M (2016) Peripartum cardiomyopathy: a systematic literature review. Acta Obstet Gynecol Scand 95:1205–1219. https://doi.org/10.1111/aogs.13005

Falagas ME, Pitsouni EI, Malietzis GA, Pappas G (2008) Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. FASEB J 22:338–342. https://doi.org/10.1096/fj.07-9492LSF

Fish M, Shaboodien G, Kraus S, Sliwa K, Seidman CE, Burke MA, Crotti L, Schwartz PJ, Mayosi BM (2016) Mutation analysis of the phospholamban gene in 315 South Africans with dilated, hypertrophic, peripartum and arrhythmogenic right ventricular cardiomyopathies. Sci Rep 6:22235. https://doi.org/10.1038/srep22235

Freeman LC, Roeder D, Mulholland RR (1979) Centrality in social networks: ii. experimental results. Soc Netw 2:119–141. https://doi.org/10.1016/0378-8733(79)90002-9

Fundikira LS, Chillo P, van Laake LW, Mutagaywa RK, Schmidt AF, Kamuhabwa A, Kwesigabo G, Asselbergs FW (2021) Risk factors and prevalence of dilated cardiomyopathy in Sub-Saharan Africa: protocol for a systematic review. JMIR Res Protoc 10:e18229. https://doi.org/10.2196/18229

Goli R, Li J, Brandimarto J, Levine LD, Riis V, McAfee Q, DePalma S, Haghighi A, Seidman JG, Seidman CE, Jacoby D, Macones G, Judge DP, Rana S, Margulies KB, Cappola TP, Alharethi R, Damp J, Hsich E, Elkayam U, Sheppard R, Alexis JD, Boehmer J, Kamiya C, Gustafsson F, Damm P, Ersbøll AS, Goland S, Hilfiker-Kleiner D, McNamara DM, The IMAC-2 and IPAC Investigators, Arany Z (2021) Genetic and phenotypic landscape of peripartum cardiomyopathy. Circulation 143:1852–1862. https://doi.org/10.1161/CIRCULATIONAHA.120.052395

Haghikia A, Podewski E, Libhaber E, Labidi S, Fischer D, Roentgen P, Tsikas D, Jordan J, Lichtinghagen R, Kaisenberg CS, Struman I, Bovy N, Sliwa K, Bauersachs J, Hilfiker-Kleiner D (2013) Phenotyping and outcome on contemporary management in a German cohort of patients with peripartum cardiomyopathy. Basic Res Cardiol 108.https://doi.org/10.1007/s00395-013-0366-9

Harakalova M, Kummeling G, Sammani A, Linschoten M, Baas AF, van der Smagt J, Doevendans PA, van Tintelen JP, Dooijes D, Mokry M, Asselbergs FW (2015) A systematic analysis of genetic dilated cardiomyopathy reveals numerous ubiquitously expressed and muscle-specific genes: A systematic analysis of

genetic DCM. Eur J Heart Fail 17:484–493. https://doi.org/10.1002/ejhf.255

Horne BD, Rasmusson KD, Alharethi R, Budge D, Brunisholz KD, Metz T, Carlquist JF, Connolly JJ, Porter TF, Lappé DL, Muhlestein JB, Silver R, Stehlik J, Park JJ, May HT, Bair TL, Anderson JL, Renlund DG, Kfoury AG (2011) Genome-wide significance and replication of the chromosome 12p11.22 locus near the PTHLH gene for peripartum cardiomyopathy. Circ Cardiovasc Genet 4:359–366. https://doi.org/10.1161/CIRCGENETICS.110.959205

Irizarry OC, Levine LD, Lewey J, Boyer T, Riis V, Elovitz MA, Arany Z (2017) Comparison of clinical characteristics and outcomes of peripartum cardiomyopathy between African American and non–African American women. JAMA Cardiol 2:1256. https://doi.org/10.1001/jamacardio.2017.3574

Isogai T, Kamiya CA (2019) Worldwide incidence of peripartum cardiomyopathy and overall maternal mortality. Int Heart J 60:503–511. https://doi.org/10.1536/ihj.18-729

Kamiya CA, Yoshimatsu J, Ikeda T (2016) Peripartum cardiomyopathy from a genetic perspective. Circ J 80:1684–1688. https://doi.org/10.1253/circj.CJ-16-0342

Katsuragi S, Tanaka H, Hasegawa J, Nakamura M, Kanayama N, Nakata M, Murakoshi T, Yoshimatsu J, Osato K, Tanaka K, Sekizawa A, Ishiwata I, Ikeda T, on behalf of the Maternal Death Exploratory Committee in Japan and Japan Association of Obstetricians and Gynecologists (2019) Analysis of preventability of hypertensive disorder in pregnancy-related maternal death using the nationwide registration system of maternal deaths in Japan. J Matern Fetal Neonatal Med 32:3420–3426. https://doi.org/10.1080/14767058.2018.1465549

Krittanawong C, Johnson KW, Rosenson RS, Wang Z, Aydar M, Baber U, Min JK, Tang WHW, Halperin JL, Narayan SM (2019) Deep learning for cardiovascular medicine: a practical primer. Eur Heart J 40:2058–2073. https://doi.org/10.1093/eurheartj/ehz056

Lannou S, Mansencal N, Couchoud C, Lassalle M, Dubourg O, Stengel B, Jacquelinet C, Charron P (2020) The public health burden of cardiomyopathies: insights from a nationwide inpatient study. J Clin Med 9:920. https://doi.org/10.3390/jcm9040920

Lewey J, Levine LD, Elovitz MA, Irizarry OC, Arany Z (2020) Importance of early diagnosis in peripartum cardiomyopathy. Hypertension 75:91–97. https://doi.org/10.1161/HYPERTENSIONAHA.119.13291

Linnenluecke MK, Marrone M, Singh AK (2020) Conducting systematic literature reviews and bibliometric analyses. Aust J Manag 45:175–194. https://doi.org/10.1177/0312896219877678

Marston S, Montgiraud C, Munster AB, Copeland O, Choi O, dos Remedios C, Messer AE, Ehler E, Knöll R (2015) OBSCN mutations associated with dilated cardiomyopathy and haploinsufficiency. PLoS ONE 10:e0138568. https://doi.org/10.1371/journal.pone.0138568

McKenna WJ, Judge DP (2021) Epidemiology of the inherited cardiomyopathies. Nat Rev Cardiol 18:22–36. https://doi.org/10.1038/s41569-020-0428-2

McKenna WJ, Maron BJ, Thiene G (2017) Classification, epidemiology, and global burden of cardiomyopathies. Circ Res 121:722–730. https://doi.org/10.1161/CIRCRESAHA.117.309711

McNamara DM, Elkayam U, Alharethi R, Damp J, Hsich E, Ewald G, Modi K, Alexis JD, Ramani GV, Semigran MJ, Haythe J, Markham DW, Marek J, Gorcsan J, Wu W-C, Lin Y, Halder I, Pisarcik J, Cooper LT, Fett JD (2015) Clinical outcomes for peripartum cardiomyopathy in North America. J Am Coll Cardiol 66:905–914. https://doi.org/10.1016/j.jacc.2015.06.1309

Mi H, Ebert D, Muruganujan A, Mills C, Albou L-P, Mushayamaha T, Thomas PD (2021) PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API. Nucleic Acids Res 49:D394–D403. https://doi.org/10.1093/nar/gkaa1106

Morales A, Painter T, Li R, Siegfried JD, Li D, Norton N, Hershberger RE (2010) Rare variant mutations in pregnancy-associated or peripartum cardiomyopathy. Circulation 121:2176–2182. https://doi.org/10.1161/CIRCULATIONAHA.109.931220

Okubo Y (1997) Bibliometric indicators and analysis of research systems: methods and examples (OECD Science, Technology and Industry Working Papers No. 1997/01), 1997. https://doi.org/10.1787/208277770603

Pearson GD, Veille JC, Rahimtoola S, Hsia J, Oakley CM, Hosenpud JD, Ansari A, Baughman KL (2020) Peripartum cardiomyopathy: National Heart, Lung, and Blood Institute and Office of Rare Diseases (National Institutes of Health) workshop recommendations and review. JAMA 283(9):1183–8. https://doi.org/10.1001/jama.283.9.1183

Peng Y, Gregorich ZR, Valeja SG, Zhang H, Cai W, Chen Y-C, Guner H, Chen AJ, Schwahn DJ, Hacker TA, Liu X, Ge Y (2014) Top-down proteomics reveals concerted reductions in myofilament and Z-disc protein phosphorylation after acute myocardial infarction. Mol Cell Proteomics 13:2752–2764. https://doi.org/10.1074/mcp.M114.040675

Penning de Vries BBL, van Smeden M, Rosendaal FR, Groenwold RHH (2020) Title, abstract, and keyword searching resulted in poor recovery of articles in systematic reviews of epidemiologic practice. J Clin Epidemiol 121:55–61. https://doi.org/10.1016/j.jclinepi.2020.01.009

Shelly S, Boaz M, Orbach H (2012) Prolactin and autoimmunity. Autoimmun Rev 11:A465–A470. https://doi.org/10.1016/j.autrev.2011.11.009

Sliwa K, Mebazaa A, Hilfiker-Kleiner D, Petrie MC, Maggioni AP, Laroche C, Regitz-Zagrosek V, Schaufelberger M, Tavazzi L, van der Meer P, Roos-Hesselink JW, Seferovic P, van Spandonck-Zwarts K, Mbakwem A, Böhm M, Mouquet F, Pieske B, Hall R, Ponikowski P, Bauersachs J (2017) Clinical characteristics of patients from the worldwide registry on peripartum cardiomyopathy (PPCM): EURObservational Research Programme in conjunction with the Heart Failure Association of the European Society of Cardiology Study Group on PPCM. Eur J Heart Fail 19:1131–1141. https://doi.org/10.1002/ejhf.780

Sliwa K, Meer P, Petrie MC, Frogoudaki A, Johnson MR, Hilfiker-Kleiner D, Hamdan R, Jackson AM, Ibrahim B, Mbakwem A, Tschöpe C, Regitz-Zagrosek V, Omerovic E, Roos-Hesselink J, Gatzoulis M, Tutarel O, Price S, Heymans S, Coats AJS, Müller C, Chioncel O, Thum T, Boer RA, Jankowska E, Ponikowski P, Lyon AR, Rosano G, Seferovic PM, Bauersachs J (2021) Risk stratification and management of women with cardiomyopathy/heart failure planning pregnancy or presenting during/after pregnancy: a position statement from the Heart Failure Association of the European Society of Cardiology Study Group on Peripartum Cardiomyopathy. Eur J Heart Fail 23:527–540. https://doi.org/10.1002/ejhf.2133

Spracklen TF, Chakafana G, Schwartz PJ, Kotta M-C, Shaboodien G, Ntusi NAB, Sliwa K (2021a) Genetics of peripartum cardiomyopathy: current knowledge, future directions and clinical implications. Genes 12:103. https://doi.org/10.3390/genes12010103

Spracklen TF, Chakafana G, Schwartz PJ, Kotta M-C, Shaboodien G, Ntusi NAB, Sliwa K (2021b) Genetics of Peripartum Cardiomyopathy: Current Knowledge, Future Directions and Clinical Implications. Genes 12:103. https://doi.org/10.3390/genes12010103

Thul PJ, Lindskog C (2018) The human protein atlas: a spatial map of the human proteome: the Human Protein Atlas. Protein Sci 27:233–244. https://doi.org/10.1002/pro.3307

Tran B, Vu G, Ha G, Vuong Q-H, Ho M-T, Vuong T-T, La V-P, Ho M-T, Nghiem K-C, Nguyen H, Latkin C, Tam W, Cheung N-M, Nguyen H-K, Ho C, Ho R (2019) Global evolution of research in artificial intelligence in health and medicine: a bibliometric study. JCM 8:360. https://doi.org/10.3390/jcm8030360

van Mook WNK, Peeters L (2005) Severe cardiac disease in pregnancy, part II: impact of congenital and acquired cardiac diseases during pregnancy. Curr Opin Crit Care 11:435–448. https://doi.org/10.1097/01.ccx.0000179806.15328.b9

van Spaendonck-Zwarts KY, van Tintelen JP, van Veldhuisen DJ, van der Werf R, Jongbloed JDH, Paulus WJ, Dooijes D, van den Berg MP (2010) Peripartum cardiomyopathy as a part of familial dilated cardiomyopathy. Circulation 121:2169–2175. https://doi.org/10.1161/CIRCULATIONAHA.109.929646

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Miklos GLG, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng Z, Francesco VD, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan W, Ge W, Gong F, Gu Z, Guan P, Heiman TJ, Higgins ME, Ji R-R, Ke Z, Ketchum KA, Lai Z, Lei Y, Li Z, Li J, Liang Y, Lin X, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue B, Sun J, Wang ZY, Wang A, Wang X, Wang J, Wei M-H, Wides R, Xiao C, Yan C, Yao A, Ye J, Zhan M, Zhang W, Zhang H, Zhao Q, Zheng L, Zhong F, Zhong W, Zhu SC, Zhao S, Gilbert D, Baumhueter S, Spier G, Carter C, Cravchik A, Woodage T, Ali F, An H, Awe A, Baldwin D, Baden H, Barnstead M, Barrow I, Beeson K, Busam D, Carver A, Center A, Cheng ML, Curry L, Danaher S, Davenport L, Desilets R, Dietz S, Dodson K, Doup L, Ferriera S, Garg N, Gluecksmann A, Hart B, Haynes J, Haynes C, Heiner C, Hladun S, Hostin D, Houck J, Howland T, Ibegwam C, Johnson J, Kalush F, Kline L, Koduru S, Love A, Mann F, May D, McCawley S, McIntosh T, McMullen I, Moy M, Moy L, Murphy B, Nelson K, Pfannkoch C, Pratts E, Puri V, Qureshi H, Reardon M, Rodriguez R, Rogers Y-H, Romblad D, Ruhfel B, Scott R, Sitter C, Smallwood M, Stewart E, Strong R, Suh E, Thomas R, Tse S, Vech C, Wang G, Wetter J, Williams S, Williams M, Windsor S, Winn-Deen E, Wolfe K, Zaveri J, Zaveri K, Abril JF, Guigo R, Campbell MJ, Sjolander KV, Karlak B, Kejariwal A, Mi H, Lazareva B, Hatton T, Narechania A, Diemer K, Muruganujan A, Guo N, Sato S, Bafna V, Istrail S, Lippert R, Schwartz R, Walenz B, Yooseph S, Allen D, Basu A, Baxendale J, Blick L, Caminha M, Carnes-Stine J, Caulk P, Chiang Y-H, Coyne M, Dahlke C, Mays AD, Dombroski M, Donnelly M, Ely D, Esparham S, Fosler C, Gire H, Glanowski S, Glasser K, Glodek A, Gorokhov M, Graham K, Gropman B, Harris M, Heil J, Henderson S, Hoover J, Jennings D, Jordan C, Jordan J, Kasha J, Kagan L, Kraft C, Levitsky A, Lewis M, Liu X, Lopez J, Ma D, Majoros W, McDaniel J, Murphy S, Newman M, Nguyen T, Nguyen N, Nodell M, Pan S, Peck J, Peterson M, Rowe W, Sanders R, Scott J, Simpson M, Smith T, Sprague A, Stockwell T, Turner R, Venter E, Wang M, Wen M, Wu D, Wu M, Xia A, Zandieh A, Zhu X (2001) The Sequence of the Human Genome. Hum Genome 291:51

Ware JS, Li J, Mazaika E, Yasso CM, DeSouza T, Cappola TP, Tsai EJ, Hilfiker-Kleiner D, Kamiya CA, Mazzarotto F, Cook SA, Halder I, Prasad SK, Pisarcik J, Hanley-Yanez K, Alharethi R, Damp J, Hsich E, Elkayam U, Sheppard R, Kealey A, Alexis J, Ramani G, Safirstein J, Boehmer J, Pauly DF, Wittstein IS, Thohan V, Zucker MJ, Liu P, Gorcsan J, McNamara DM, Seidman CE, Seidman JG, Arany Z (2016) Shared genetic predisposition in peripartum and dilated cardiomyopathies. N Engl J Med 374:233–241. https://doi.org/10.1056/NEJMoa1505517

Wu M, Zhang Y, Grosser M, Tipper S, Venter D, Lin H, Lu J (2021) Profiling COVID-19 Genetic Research: A Data-Driven Study Utilizing Intelligent Bibliometrics. Front Res Metr Anal 6. https://doi.org/10.3389/frma.2021.683212

Zhang Y, Zhang G, Zhu D, Lu J (2017) Scientific evolutionary pathways: Identifying and visualizing relationships for scientific topics. J Assoc Inf Sci Technol 68:1925–1939. https://doi.org/10.1002/asi.23814

Zhang Y, Wu M, Lin H, Tipper S, Grosser M, Zhang G, Lu J (2020) Framework of computational intelligence-enhanced knowledge base construction: methodology and a case of gene-related cardiovascular disease. Int J Comput Intell Syst 13:1109–1119. https://doi.org/10.2991/ijcis.d.200728.001