The Journal of Molecular Diagnostics, Vol. 27, No. 4, April 2025

ELSEVIER

the Journal of Molecular Diagnostics
jmdjournal.org

Biomimetic Digital Twins and Multiomics

Applications to Rheumatoid Arthritis and the Potential Reclassification of Variants of Unknown Clinical Significance

William G. Kearns,*† Joe Glick,‡ Lawrence Baisch,‡ Andrew Benner,* Dalton Brough,* Luke Du,* Chandra Germain,* Laura Kearns,*† and Georgios Stamoulis§

From Genzeva, Rockville, Maryland; LumaGene,† Rockville, Maryland; RYLTI BioPharma,‡ Hauppauge, New York; and Qiagen Digital Insights,§ Redwood City, Californ*

Accepted for publication December 26, 2024.

Address correspondence to William G. Kearns, Ph.D., Genzeva, 9430 Key West Ave., Ste. 130, Rockville, MD 20850.
E-mail: wgkearns@genzeva.com.

The National Academies of Sciences, Engineering, and Medicine issued a report on December 15, 2023, "Foundational Research Gaps and Future Directions for Digital Twins." This described the importance of using biomimetic digital twins and multiomics in research. These were incorporated in the current analysis of patients with rheumatoid arthritis (RA). Exome sequencing, genotype-phenotype ranking, and biomimetic digital twin analysis were used to identify five pathogenic and one likely pathogenic DNA variants in patient samples analyzed, which were absent from controls. The variants identified in these genes, *P2RX7*, *HTRA2*, *PTPN22*, *FLG*, *CD46*, and *EIF4G1*, play a role in the development of RA. Additionally, 3172 variants of unknown clinical significance (VUSs) were identified in patient samples, which were absent from controls. All VUSs appeared to be associated with RA. Hidden or dark data were identified from six genes. These genes, often found in patient samples, included *HIF1A*, *HLA-DOA*, *PTGER3*, *HIPK3*, *TGFBR3*, and *HIF1A-AS3*. VUSs identified in genes *HIF1A*, *HLA-DOA*, *PTGER3*, and *HIPK3* were directly related to the pathogenesis of RA, whereas VUSs identified in genes *TGFBR3* and *HIF1A-AS3* were indirectly related. The current results suggest that biomimetic digital twins and multiomics can provide further insight into the development of RA. This may also potentially help with the process of reclassifying VUSs. The reclassification of VUSs will play a critical role in complex molecular diagnostics and drug development. *(J Mol Diagn 2025, 27: 256—269; https://doi.org/10.1016/j.jmoldx.2024.12.012)*

Modeling real-world complexity has long been the goal of biomedical methods and software, but it persists in being a major challenge because of the astronomical combinatorial possibilities of biological systems, which are multidimensional and multiscale. Modeling the interactions within molecular genomic ecosystems and how they interact with human biology has long been the driver of biomedical research and technological advancement.[1,2]

The futility of trying to model real-world complexity by scaling data volume and processing power has been identified and communicated for nearly two decades but has not been adequately assimilated.

Artificial intelligence (AI) and other technologies such as machine learning (ML), neural networks, and large language models hold tremendous promise for driving advances in biomedical research. However, along with their benefits, they also have limitations.

Within a biomimetic digital twin method, the raw data undergo no cleansing or normalization. The classification is driven by expert knowledge graphs. The software models the complex relationships across diverse small and large data using the expertise graphs to compute relevance. The outputs identify all the potential relationship scenarios between data points. The researchers then review the potential evidence and draw conclusions. Often, the reports reveal

Supported by the Genzeva (W.G.K.), LumaGene (W.G.K.), and Ryailiti Research Programs (J.G.).

https://doi.org/10.1016/j.jmoldx.2024.12.012

hidden or dark data, relationships that were unknown to the researchers beforehand.

The National Academies of Sciences, Engineering, and Medicine (NAS) published a Physics of Life Report,[3] which concluded, "An important lesson from the long and complex history of neural networks and AI is that revolutionary technology can be based on ideas and principles drawn from an understanding of life, rather than on direct harnessing of life's mechanisms or hardware."[3,p.242]

To address these issues and to provide guidance to the biomedical community, the NAS, sponsored by the NIH, the National Science Foundation, the Defense Advanced Research Projects Agency, and the Department of Energy, began advocating research into the use of biomimetic digital twin technology to more effectively model multidimensional and multiscale biological complexity.[2]

On December 15, 2023, the NAS released a 164-page report, "Foundational Research Gaps and Future Directions for Digital Twins."[4] The report stated: "Across multiple domains of science, engineering, and medicine, excitement is growing about the potential of digital twins to transform scientific research, industrial practices, and many aspects of daily life. A digital twin combines computational models with a physical counterpart to create a system that is dynamically updated through bidirectional data flows as conditions change. Going beyond traditional simulation and modeling, digital twins could enable improved precision medicine and healthcare by more clearly understanding the pathophysiology of disease. This report identified the foundational research and resources needed to support the development of digital twin technologies. The report presents critical future research priorities and an interdisciplinary research agenda for the field, including how federal agencies and researchers across domains can best collaborate."[4,p.114]

Although this report advocated for the use of AI, ML, and neural networks in many forms of biomedical research, it also stated that digital twins can further assist in addressing the NAS-identified research gaps. This includes the conclusion that new theories and methods are required to address the multidimensional, multiscale characteristics of problems in modeling and advanced analytics in general, and in biomedicine in particular.

However, digital twins are not designed to think like a human brain. Biomimetic digital twins are designed to add human thinking to digital twin technologies and incorporate that human expertise into analytical computations. Multiomics and biomimetic digital twins technology significantly assists in filling the research gaps identified by the NAS report for biomedical research.

In harmony with the NAS guidance, Kearns et al[5] (2024) incorporated a biomimetic digital twin ecosystem into advanced multiomics experimental protocols. It used a biomimetic knowledge engineering method to generate an ecosystem of digital twins that implement real-world reasoning principles and analyzed data that are raw and in their original state, meaning that no cleansing or normalization was performed to remove outliers and/or hide relationships and impacts within data sets. The use of this method both leveraged and used dark or hidden data and enabled unexpected discovery. It provided evidence for a potential biomarker for a less invasive diagnostic for endometriosis, and identified a potential chromosomal hotspot associated with the pathogenesis of endometriosis. Thus, multiomics and biomimetic digital twins can help researchers more clearly define the molecular mechanism of disease.

Key Points

- To our knowledge, this is the first submitted research article following the National Academies of Sciences, Engineering, and Medicine recommendations, released on December 15, 2023, that digital twins should be incorporated in biomedical research to close research gaps.
- Herein, biomimetic digital twins were incorporated into a comprehensive multiomics platform to potentially clarify the pathogenesis of a complex disorder, rheumatoid arthritis. This was accomplished by identifying dark or hidden data—complex relationships that are not visible in bioinformatics platforms—as either directly or indirectly related to the development of rheumatoid arthritis.
- The current results suggest that biomimetic digital twins and a comprehensive multiomics platform can help in the process of reclassifying variants of unknown clinical significance (VUSs).
- The reclassification of VUSs would play a critical role in complex diagnostics and drug development.

The current study focused on the molecular mechanisms of rheumatoid arthritis (RA).[6–9] RA is a multifactorial autoimmune disease of unknown etiology, primarily affecting the joints; extra-articular manifestations can occur. RA causes joint inflammation, which in severe cases may result in permanent joint damage and disability. Additionally, RA may affect other organs, including the lungs, heart, blood vessels, skin, and eyes. RA affects approximately 1 of every 200 adults worldwide and occurs two to three times more frequently in women than men. It can affect people of any age, but peak onset is from age 50 to 59 years.

Most epidemiologic studies in RA have been conducted in western countries, showing an RA prevalence in the range of 0.5 to 1.0% in the United States. The cumulative lifetime risk of developing adult-onset RA has been estimated at 3.6% for women and 1.7% for men.[10]

RA has a strong genetic component.[10] Twin studies have estimated the heritability of RA to be approximately 60%. This number is observed in anti-cyclic citrullinated peptide antibody (ACPA)−positive patients. These patients have a more severe subset of RA, with more severe joint destruction and a higher mortality rate. ACPA positivity is also associated with older age, female sex, smoking, joint complaints, and first-degree relatives with RA.

The disease concordance of identical twins is only 12% to 15%, indicating that environmental factors also play an important role in susceptibility.

Genotype-phenotype relationships have identified >6000 genes with some potential association to the pathophysiology of RA.[11–13]

Over 100 loci have been identified across genomes harboring RA susceptibility variants by genome-wide association studies, with fine mapping, candidate gene approaches, and a meta-analysis of genome-wide association studies involving >100,000 individuals.[14,15]

RA is a complex multifactorial disease with both genetic and environmental risk factors contributing to it, and multiple risk factors may be required before reaching the threshold at which RA is triggered.

In this study, exome sequencing, DNA variant phenotype-driven ranking analysis, biomimetic digital twin analysis, GeneCards (*https://www.genecards.org*, accessed January 2024), and VarElect (LifeMap Sciences, Hong Kong) were used to identify dark or hidden data associated with the molecular profile of RA, a complex multifactorial disorder. It provided additional results demonstrating that biomimetic digital twins and comprehensive multiomics can play a role in the clarification in the pathophysiology of a complex genetic disorder. Furthermore, it demonstrated the potential role of using this platform in identifying variants of unknown clinical significance (VUSs) potentially associated with the development of rheumatoid arthritis and suggest a way to potentially begin to reclassify VUSs as pathogenic, likely pathogenic, benign, or likely benign.

Materials and Methods

Patient Population

All patient samples analyzed in this study were obtained from immunodeficiency exome clinical tests ordered by physicians with a signed informed consent. The patient ages ranged from 65 to 72 years, 19 were White, 4 were African American, and 2 were Asian. Thirteen samples were from females, and 12 were from males. All controls were from patients aged >60 years, with no diagnosis of RA or related disorders, and with no identified comorbidities.

Next-Generation Sequencing

Experimental protocol: whole-exome next-generation sequencing was performed on each sample to determine the presence or absence of known pathogenic, or likely pathogenic, mutations and VUSs associated with RA.

Whole-genome amplified DNA (50 ng) from each sample was used as input for library preparation (Thermo Fisher Scientific, Waltham, MA). The library preparation was done using xGen DNA Library Prep EZ UNI (Integrated DNA Technologies, Coralville, IA). The DNA sample underwent enzymatic preparation to produce fragment sizes of approximately 200 bp. This was followed by ligation using full-length adapters. The samples then underwent an AMPpure bead (Beckman Coulter, Sharon, IL) cleanup and were washed. A PCR amplification was then performed, followed by a second AMPure bead cleanup. The samples were then sized (4200 TapeStation; Agilent Technologies, Santa Clara, CA) and quantitated (Qubit 4 Fluorometer; Fisher Scientific, Waltham, MA). Samples were pooled with no more than 12 samples per pool, and a 16-hour hybridization was preformed using xGen Exome Hyb Panel v2 (Integrated DNA Technologies).

A bead capture (Bait-Capture) and a set of post-hybridization washes were performed using an xGen Hybridization and Wash Kit (Integrated DNA Technologies). A post-hybridization amplification was then done using xGen Library Amplification Primers (Integrated DNA Technologies), followed by an AMPure bead cleanup. The pools were sized and quantitated once more. The pools were normalized and pooled into a single pool.

The pooled libraries were then denatured and loaded onto a NovaSeq 6000 (Illumina, San Diego, CA) and sequenced using a NovaSeq 6000 S1 Reagent Kit v1.5 (Illumina). The libraries bind to grafted oligos on the flow cell and then hybridize and bridge on their specific oligo and undergo multiple cycles of amplification. This forms clusters using an ExAmp technology. Then, the clusters undergo two-channel sequencing by synthesis chemistry.

Validation

The experimental protocol included a NovaSeq 6000 for short-read next-generation sequencing, the Illumina Dragen Germline pathway for secondary analysis, and Qiagen's Clinical Insight (Redwood City, CA) for tertiary analysis. First, whole-exome sequencing was comprehensively validated against National Institute of Standards and Technology reference/validation samples. TAccuracy, sensitivity, specificity, positive predictive value, negative predictive value, positive percentage agreement, and precision (inter- and intra-) assays were performed to complete the validation process.

The authors also participate in the College of Pathologists surveys. Blinded DNA sequencing to previously known samples was also performed to ensure the accuracy of the results.

Required passing quality control metrics for each sample sequenced were as follows: Total_input_reads: >49,000,000; Number_of_duplicate_marked_reads_pct: <10%; Uniformity_of_coverage_pct_gt_02mean_over_target_region: >95%; Average_alignment_coverage_over_target_region: >85%; and Pct_of_target_region_with_coverage_20x_inf: >95%.

Short-read sequencing (approximately 350 bp) was the best modality to use for this study as long-range sequencing (approximately 3000 to 5000 bp) is more

relevant to identify structural variants. Furthermore, the current standard of care for clinical next-generation sequencing testing uses short-read sequencing.

Secondary Analysis

On sequencing, the Dragen Platform (Illumina) was used for secondary analysis. This enrichment is an accurate and efficient end-to-end (FASTQ to VCF) secondary analysis solution for whole-exome data. This app takes input files in FASTQ, BAM, and CRAM format. Files may be decompressed, go through map/align/sort, and go through variant calling using Qiagen's Clinical Insight.

Tertiary Analysis

For tertiary analysis, all variants and all phenotype ranked variants were downloaded using Qiagen's Clinical Insight (Qiagen Digital Insights). Qiagen's Clinical Insights Interpret is a clinical decision support software that accelerates variant interpretation and reporting of Mendelian, hereditary, rare disease, complex disorders, and oncology next-generation sequencing tests at scale. Qiagen's Clinical Insights Interpret is powered by Qiagen Knowledge Base, the biggest manual curated knowledgebase, with insights about symptoms, phenotypes, and gene-disease associations, biomedical databases, such as Human Gene Mutation Database and Catalogue of Somatic Mutations in Cancer, medical guidelines, and a wide variety of different bibliography content sources that are clinically relevant and are manually curated daily. Qiagen's Clinical Insight Interpret computes and combines all the relevant information related to the variant of interest and distributes the relevant biological context. Qiagen's Clinical Insight also offers the possibility of phenotype-driven analysis, where the user can submit phenotypes or symptoms of suspected disease or disease under investigation along with the .vcf file of the sample. On the basis of this information, Qiagen's Clinical Insights Interpret phenotype-driven ranking algorithm estimates and ranks genomic variants based on the probability of being the causative one for the disease, symptoms, or the phenotypes under investigation by taking into account multiple variables, such as zygosity, predicted pathogenicity of variant, mode of inheritance, Combined Annotation Dependent Depletion (CADD) score, and more variant-centric variables, as well as all the curated molecular insights from the Qiagen knowledge base.

Comparison Between Standard Artificial Intelligence and a Biomimetic Digital Twin Analysis

Biomimetic Digital Twin Architecture Overview

- Human expertise graphs
- Model and ecosystem design
- Real-world data approach
- Dark data discovery
- Transparency

Human Expertise Graphs

Experts use qualitative reasoning for problem analysis. Therefore, the biomimetic digital twin ecosystem must include a qualitative meta ontology with domains that can be populated and mapped independently by the subject matter experts. Here is a high-level example (Figure 1).

Experts map relevant attributes in the provided data sources to their own qualitative models of the domains or to industry standard ontologies.

Model and Ecosystem Design

Models are scoped around known behaviors and designed by imitating (twinning) the understood structures, systems, and scenarios of the modeled behaviors. Emerging behaviors are not predictions, but evidence to be considered by experts (Figure 2).

Ecosystem Architecture

- Each twin models a discrete component of the analytical scope of the ecosystem.
- Internal properties and behaviors must be modeled to a level of sufficient comprehensiveness to enable the reactions that are required for the ecosystem to reflect the real world to the scope of its design.
- Each twin can initiate an interaction with others or respond as prompted.
- Mitigation of bias is achieved by:
  - Independent design of each twin
  - Abstract knowledge graphs populated without defining specific problems or events
  - Autonomous interactions between the twins

Real-World Data Approach

A data lake is populated with the required small/wide data sources, which could be small data sets, such as extracts from patient records, or outputs of larger systems, such as bioinformatics platforms. All tables are in their native schema without normalization or cleansing—real-world data.

Contextualization is the primary method of interpreting data and assessing the relevance of evidence to a defined problem. Big data are more likely to pose challenges rather than help. A recent article titled "The Limits of Data," from the National Academy of Sciences, concluded "Data is powerful because it's universal. The cost is context." (*https://issues.org/limits-of-data-nguyen*, last accessed November 1, 2024).

Hidden or Dark Data Discovery

The relevance computation engine leverages the combined expertise graphs to identify multiscale and multidimensional relationships across the data sets in the lake. This is an

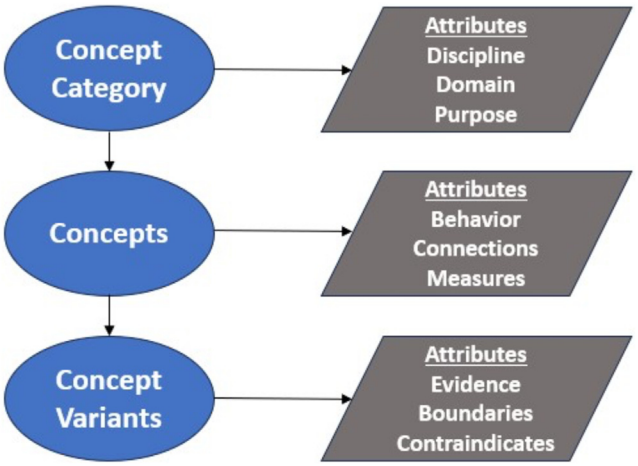

Figure 1 The use of concept categories, concepts, and concept variants to generate human expertise graphs.

engineering-level view of dark or hidden data discovery (Figure 3).

In this case, multifactor correlations between pathogenic variants (relevant), VUSs (other), and knowledge graphs populated by the researchers produced the reported findings.

The findings cannot be compared with AI outputs using the same inputs because standard AI: i) requires large training data sets; ii) statistically computes predictions rather than discovering contextual relationships, so they cannot deliver real-world data but frequently produce hallucinations; and iii) uses black box algorithms that provide no explanation for deriving the outputs.

Transparency

The biomimetic digital twin ecosystem is NOT a black box application because:

- The process of computing relevance using expert mappings is transparent and does not perform any data transformation.
- The outputs are not predictions but discovered relationships with the associated evidence.
- The outputs identify the source files and attributes for each value that is presented as part of the evidence, so that everything is traceable.

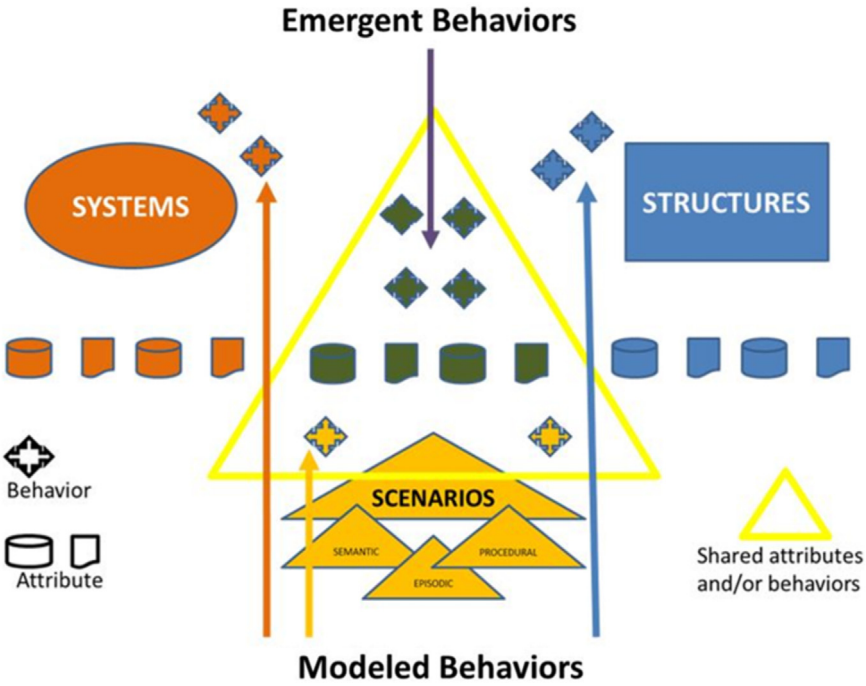

Figure 2 The use of systems, structures, modeled behaviors, and emergent behaviors as a model and ecosystem design.

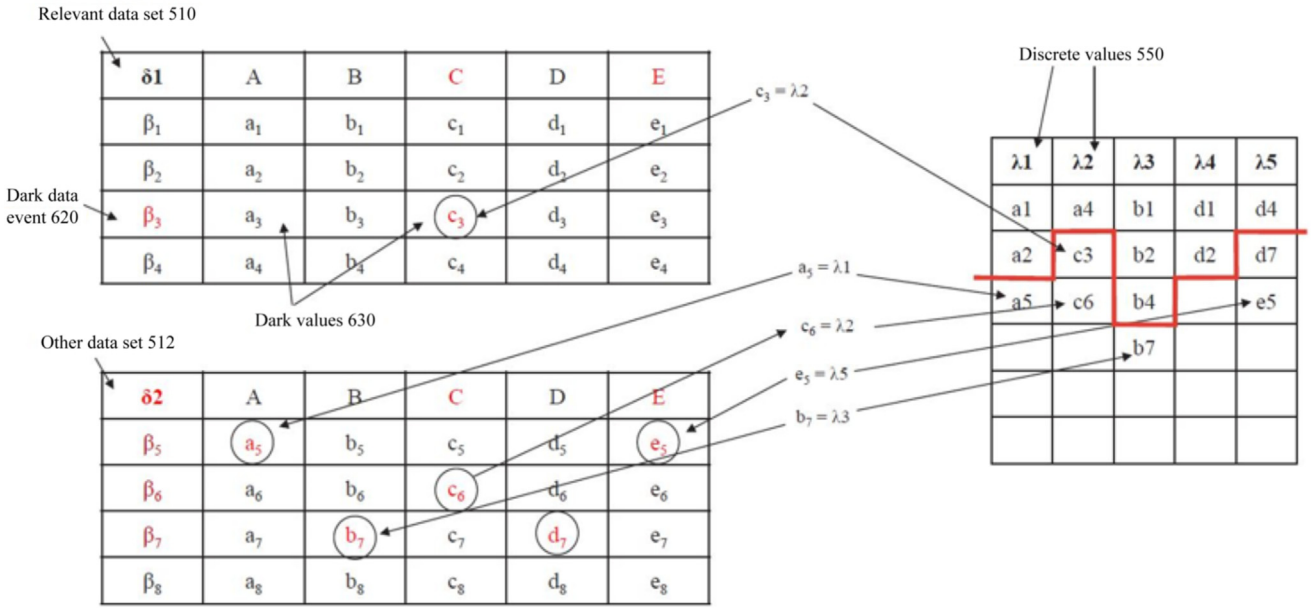

Figure 3 This illustrates hidden specific correlations (dark data) discovered by the biomimetic digital twin engine. The figure illustrates unknown correlations between factors in disparate data sets. **Red text** and **red line** indicate the most prevalent correlations.

- The only variable may be differences in experts' views on the significance of the evidence.

Knowledge Engineering Using a Biomimetic Engine

- Each twin models a discrete component of the analytical scope of the ecosystem.[2–4,16–19]
- Internal properties and behaviors must be modeled to a level of sufficient comprehensiveness to enable the reactions that are required for the ecosystem to reflect the real world to the scope of its design.
- Each twin can initiate an interaction with others or respond as prompted.
- Mitigation of bias is achieved by:
  - Independent design of each twin
  - Abstract knowledge graphs populated without defining specific problems or events
  - Autonomous interactions between the twins

This real-world reasoning approach enables the construction of models that integrate highly diverse elements and information sources to enable exploration and discovery to a scope that traditional information architecture cannot accommodate.

Systems Thinking and Real-World Reasoning

The NAS also recommends addressing complexity using systems thinking. Key observations are as follows:

- Bottom-up, mechanistic, linear approaches to understanding macro-level behavior are limited when considering complex systems.
- Bottom-up, reductionist hypotheses and approaches can lead to a proliferation of parameters; this challenge can potentially be addressed by applying top-down, system-level principles.
- Systems thinking can be used to predict macroscopic phenomena while bypassing the need to explicitly unmask all the quantitative dynamics operating at the microscopic level.

Although all knowledge engineering efforts seek to incorporate elements of cognitive science, a key aspect of this innovation strategy is the driving role of a cognitive method, which is enabled by biomimetic information architectures. Brain processes are systemic and leverage what neuroscientists label plasticity and sparsity.

- Plasticity is the ability to engage diverse combinations of neurons and synapses by relevance to the purpose of the analysis, and to dynamically adapt internal functional architectures.
- Sparsity is the ability to identify the minimum data required. The brain can respond to situations that are simultaneously new on multiple dimensions and can even categorize one data point.

The neuronal and synaptic architecture of the brain is an ecosystem, which, according to the NAS, contains 100 trillion neurons. Systemic architecture, plasticity, and sparsity are core to biological learning, but are NOT like ML algorithms. The biomimetic technologies that enable elements of real-world reasoning are as follows:

- Expertise graphs
- Neural system dynamics digital twins

Principles of plasticity and sparsity can be imitated by implementing qualitative expertise graphs and leveraging them for contextual selection of data and methods from the in-memory model library. Unlike the deterministic methods by which traditional application engineering is limited, systemic modeling requires the coexistence of chaotic and stochastic model elements, as well as their ability to dynamically interact with the deterministic elements.

For several years, AI has been looked to as the leading pathway to genetic understanding and drug development. However, deep learning and natural language processing have the three key challenges listed in *Hidden or Dark Data Discovery* that are addressed by the biomimetic digital twin ecosystem method presented in this article.

Multiomics and Biomimetic Digital Twin Ecosystem Process Tailored for the Analysis

i) Each patient DNA sample and controls underwent exome sequencing, secondary analysis, and tertiary analysis (Figure 4). All DNA variants and phenotype ranked variants were exported to the digital twin ecosystem's data lake.

ii) Expert knowledge graphs were produced listing all previously reported DNA variants potentially associated with the pathophysiology of RA and were exported to the digital twin ecosystem's data lake.

iii) Expert knowledge graphs were produced from each patient medical record and exported to the digital twin ecosystem's data lake.

iv) The digital twin ecosystem's biomimetic engine then combined all data downloaded from Clinical Insight, including *in silico* calculations, phenotype ranked references, and multifactor correlations to the generated knowledge graphs, and produced a list of gene variants classified as VUSs potentially associated with the pathophysiology of RA.

v) The digital twin ecosystem's biomimetic engine ranked all VUSs according to the number of times that they were present in patient samples but absent from controls.

vi) The digital twin ecosystem's biomimetic engine's output pinpointed six genes, with DNA variants classified as VUSs, and phenotypes potentially associated with the pathophysiology of RA.

Figure 4 Multiomics and biomimetic digital twin ecosystem experimental protocol design tailored for the analysis. QCI, Qiagen's Clinical Insight; RA, rheumatoid arthritis; VUS, variant of uncertain significance.

vii) The digital twin output data were uploaded into GeneCards and VarElect to identify genotype/phenotype relationships associated with pathophysiology of RA.
   a. GeneCards is a searchable, integrative database that provides comprehensive information on all annotated and predicted human genes. The knowledgebase automatically integrates gene-centric data from approximately 200 web sources, including genomic, transcriptomic, proteomic, genetic, clinical, and functional information.[11,12]
   b. VarElect is a comprehensive phenotype-dependent DNA variant/gene prioritizer that can identify causal DNA variants with phenotypes. VarElect provides search and scoring capabilities, proficiently matching DNA variant-containing genes to submitted disease/symptom/phenotype keywords. The VarElect algorithm infers direct as well as indirect links between genes and phenotypes.[13]

viii) The digital twin ecosystem's biomimetic engine does not make recommendations or draw conclusions, but rather provides researchers with evidence for consideration that is not visible to standard AI or traditional bioinformatics platforms or approaches.

Statistical Analysis

Although statistical methods may ordinarily be applied to the data at this stage in the analysis, this method enables researchers to discover real-world evidence that they cannot find using standard research software, including ML/AI tools. Assessing the statistical significance of the evidence, if desired, can be performed, but the calculations depend on the researcher's hypotheses in combination with other available evidence. The use of *P* values and associated methods is not without controversy.[20] The current approach delivers the results and the supporting evidence, and adding a statistical component to the outcome could reduce the clarity of the results and possibly add bias.

Results

Phenotype Ranking of DNA Variants

All DNA variants were downloaded from Qiagen's Clinical Insight from each patient's sample and controls.

The Human Phenotype Ontology term HP:0001370 (*https://hpo.jax.org/browse/term/HP:0001370*, last accessed November 1, 2024) was used for RA for phenotype ranking analysis

Five pathogenic and one likely pathogenic DNA variant were identified in 8 of 25 patient samples analyzed. All these mutations are associated with RA. These include genes *P2RX7*, *HTRA2*, *PTPN22* (likely pathogenic), *FLG*, *CD46*, and *EIF4G1*. No pathogenic or likely pathogenic DNA variant was identified in any controls (Table 1).

P2RX7: This gene is a highly expressed receptor on immune cells, triggering the release of cytokines and regulating autoimmune responses. The synthesis of proinflammatory cytokines and apoptosis of lymphoid cells can be induced through *P2X7*. These results suggest a possible involvement of *P2X7* in the pathogenesis of inflammatory autoimmune diseases and its role in the development of RA.[11–13]

HTRA2: This is a serine peptidase that plays a significant role in collagen-induced RA. *HTRA2* modulates inflammatory responses by controlling *TRAF2* stability in collagen-induced RA.[11–13]

PTPN22: This gene encodes a member of the nonreceptor class 4 subfamily of the protein-tyrosine phosphatase family. The encoded protein is a lymphoid-specific intracellular phosphatase that associates with the molecular adapter protein casitas B-lineage lymphoma (CBL) and may be involved in regulating CBL function in the T-cell receptor signaling pathway. DNA variants in this gene may be associated with a range of autoimmune disorders, including type 1 diabetes, RA, systemic lupus erythematosus, and Graves' disease.[11–13]

FLG: Antikeratin antibodies and the antiperinuclear factor are the most specific serological markers of RA. They are largely the same autoantibodies that recognizes human epidermal filaggrins and profilaggrin-related proteins of buccal epithelial cells (collectively referred to as profilaggrin).[11–13]

CD46: The protein encoded by this gene is a type I membrane protein and is a regulatory part of the complement system. *CD46* acts as a cofactor for complement factor I, a serine protease that protects autologous cells against complement-mediated injury by cleaving C3b and C4b deposited on host tissue. *CD46* acts as a costimulatory factor for T cells, which induces the differentiation of $CD4^+$ into T-regulatory 1 cells. T-regulatory 1 cells suppress the immune system.[11–13]

EIF4G1: The protein encoded by this gene is a component of the multisubunit protein complex eukaryotic translation initiation factor 4F (EIF4F). This complex facilitates the recruitment of mRNA to the ribosome, which is a rate-limiting step during the initiation phase of protein synthesis. The recognition of the mRNA cap and the ATP-dependent unwinding of 5′-terminal secondary structure are catalyzed by factors in this complex. The subunit encoded by this gene is a large scaffolding protein that contains binding sites for other members of the EIF4F complex. A domain at its N-terminus can also interact with the poly(A)-binding protein, which may mediate the circularization of mRNA during translation. Pathogenic DNA variants within this gene dysregulate the recruitment of mRNA to ribosomes and are associated with pathophysiology of RA.[11–13]

Table 1 Gene, Transcript Variants, Protein Variants, Translation Impact, and CADD Scores for Six Genes Related to the Pathophysiology of Rheumatoid Arthritis

Gene symbol	Transcript variant	Protein variant	Translation impact	CADD score	Sample
HIF1A	c.2038C>G; c.2107C>G; c.2035C>G; n.213+9755G>C	p.Q680E; p.Q703E; p.Q679E	Missense	16.54	10,461
HIF1A	c.35+2003delT; c.-9delT			3.912	10,430
HIF1A	c.1256C>T; c.1253C>T; c.1325C>T; n.213+12795G>A	p.T419I; p.T419I; p.T442I	Missense	20.3	10,444
HIF1A	c.1256C>T; c.1253C>T; c.1325C>T; n.213+12795G>A	p.T419I; p.T419I; p.T442I	Missense	20.3	10,454
HIF1A	c.151G>C; n.214-3477C>G; c.220G>C; c.148G>C	p.V74L; p.V51L; p.V50L	Missense	23.4	10,431
HIF1A	c.44delT; c.35+2055delT	p.L15*	Frameshift	20.2	10,500
HIF1A	c.44delT; c.35+2055delT	p.L15*	Frameshift	20.2	10,500
HIF1A-AS3	c.2038C>G; c.2107C>G; c.2035C>G; n.213+9755G>C	p.Q680E; p.Q703E; p.Q679E	Missense	16.54	10,461
HIF1A-AS3	c.1256C>T; c.1253C>T; c.1325C>T; n.213+12795G>A	p.T419I; p.T419I; p.T442I	Missense	20.3	10,444
HIF1A-AS3	c.1256C>T; c.1253C>T; c.1325C>T; n.213+12795G>A	p.T419I; p.T419I; p.T442I	Missense	20.3	10,454
HIF1A-AS3	c.151G>C; n.214-3477C>G; c.220G>C; c.148G>C	p.V74L; p.V51L; p.V50L	Missense	23.4	10,431
HIF1A-AS3	c.2355G>A; n.213+3929C>T; c.2424G>A; c.*17G>A; c.2352G>A	p.G784G; p.G785G; p.G808G	Synonymous	<10	10,500
HIPK3	c.509G>A	p.G170E	Missense	20.7	10,447
HIPK3	c.732A>G	p.I244M	Missense	23.3	10,491
HIPK3	c.3511C>T; c.3448C>T	p.R1171C; p.R1150C	Missense	28.2	10,436
HIPK3	c.1499G>A	p.S500N	Missense	20.7	10,500
HLA-DOA	c.313C>T	p.R105C	Missense	23.2	10,447
HLA-DOA	c.313C>T	p.R105C	Missense	23.2	10,455
HLA-DOA	c.108C>T	p.P36P	Synonymous	<10	10,465
HLA-DOA	c.3G>A	p.M1I	Start loss	24.7	10,466
PTGER3	n.1316+59326delC; c.1185delC; c.*104delC; c.1104+784delC; n.1343+784delC; c.1077+59326delC; c.*23+784delC	p.N395fs*9	Frameshift	19.64	10,491
PTGER3	c.1105T>C; c.1077+59246T>C; c.*24&>C; n.1316+59246T>C; c.1104+704T>C; c.*23+704T>C; n.1343+704T>C	p.L369L	Synonymous	<10	10,446
PTGER3	n.1317-20553C>T; c.1124C>T; c.1078-20553C>T; n.1316+37963C>T; c.1077+37963C>T	p.P375L	Missense	20.3	10,446
PTGER3	n.1316+59326delC; c.1185delC; c.*104delC; c.1104+784delC; n.1343+784delC; c.1077+59326delC; c.*23+784delC	p.N395fs*9	Frameshift	19.64	10,440
TGFBR3	c.2329C>T; c. 2326C>T; n.2813C>T	p.P777S; p.P776S	Missense	22.8	10,447
TGFBR3	c.2365A>T; n.2852A>T; c.2368A>T	p.I790F; p.I789F	Missense	29.8	10,451
TGFBR3	c.886G>T; n.1370G>T	p.A296S	Missense	22.9	10,446
TGFBR3	n.442A>G; c.55A>G	p.T19A	Missense	<10	10,454
TGFBR3	c.464A>G; n.948A>G	p.H155R	Missense	19.22	10,484

Combining Phenotype Ranking and Biomimetic Digital Twin Analysis

Next, all genotype-phenotype ranked variants were downloaded using specific key terms that described the phenotype of RA using a phenotype-driven ranking filter (Qiagen's Clinical Insight Interpret) for each patient sample. The data were then exported into the biomimetic digital twin ecosystem for analysis. It identified 3172 VUSs in patient samples analyzed, but not in controls.

Table 2 Genes Associated with a Direct Relationship to the Pathophysiology of Rheumatoid Arthritis

Gene symbol	Description	Category
HLA-DOA	Major histocompatibility complex, class II, DO α	Protein coding
HIF1A	Hypoxia-inducible factor 1 subunit α	Protein coding
PTGER3	Prostaglandin E receptor 3	Protein coding
HIPK3	Homeodomain-interacting protein kinase 3	Protein coding

Run from VarElect (Copyright © LifeMap Sciences, Inc.), used with permission.

Hidden or dark data for DNA variants were identified in six genes classified as VUSs in patient samples. The genes often found in patient samples included *HIF1A*, *HLA-DOA*, *PTGER3*, *HIPK3*, *TGFBR3*, and *HIF1A-AS3* (Tables 2 and 3).

HIF1A: This gene encodes the α subunit of transcription factor hypoxia-inducible factor-1 (HIF-1), which is a heterodimer composed of an α and a β subunit. HIF-1 functions as a master regulator of cellular and systemic homeostatic response to hypoxia by activating transcription of many genes, including those involved in energy metabolism, angiogenesis, apoptosis, and other genes whose protein products increase oxygen delivery or facilitate metabolic adaptation to hypoxia.[11–13]

Eighteen VUSs and 12 different proteins within the *HIF1A* gene were identified in patients analyzed. All but one was classified as a missense mutation.

HLA-DOA: *HLA-DOA* is a protein-coding gene that belongs to the human leukocyte antigen (HLA) class II α chain paralogues. It is a non-classic HLA gene that forms a heterodimer with *HLA-DOB*. The heterodimer, *HLA-DOA*, is found in lysosomes in B cells and regulates HLA-DM–mediated peptide loading on major histocompatibility complex class II molecules. One study identified an independent risk of a synonymous mutation at *HLA-DOA* on ACPA-positive RA risk.[11–13]

Three VUSs and three different proteins were identified within the *HLA-DOA* gene in patients analyzed.

One was a missense mutation, one was a synonymous variant, and one was a start-loss mutation.

PTGER3: This is a receptor for prostaglandin E2 (PGE2). The activity of this receptor can couple to both the inhibition of adenylate cyclase mediated by G(i) proteins and to an elevation of intracellular calcium. Prostanoid receptors are activated by the endogenous ligands prostaglandin (PG) D2, PGE2, PGF2α, PGH2, prostacyclin (PGI2), and thromboxane A2. Cyclooxygenase converts arachidonic acid to PGH2, from which other prostaglandins are synthesized. PGE2 is induced with IL-1, which also enhances the production of parathyroid hormone–related protein. The induction of PGE2 by IL-1α appears to be an important component of the parathyroid hormone–related protein production of the inflammatory process in synovial tissues from patients with RA.[11–13]

Twenty-one VUSs were identified that encoded four different proteins within the *PTGER3* gene in patients analyzed. Two were frameshift variants, one was a synonymous mutation, and one was a missense variant.

HIPK3: This gene enables protein serine/threonine kinase activity, is involved in mRNA transcription, provides negative regulation of apoptosis, and aids in protein phosphorylation. DNA variants within this gene appear to play a role in the development of RA.[11–13]

Five VUSs were identified that encoded five different proteins within the *HIPK3* gene in patients analyzed. All proteins were classified as missense variants.

TGFBR3: This locus encodes the transforming growth factor (TGF)-β type III receptor. The encoded receptor is a membrane proteoglycan that often functions as a coreceptor with other TGF-β receptor superfamily members. Ectodomain shedding produces soluble transforming growth factor beta receptor III (TGFBR3), which may inhibit transforming growth factor beta protein (TGFB) signaling. Variants with this gene likely play an indirect role in the pathophysiology of RA.[11–13]

Twelve VUSs were identified encoding seven proteins within the *TGFBR3* gene in patients analyzed. All variants were classified as missense mutations.

Table 3 Genes Associated with an Indirect Relationship to the Pathophysiology of Rheumatoid Arthritis

Gene symbol	Pathway	Description	Category
TGFBR3	TNF	Tumor necrosis factor	Protein coding
TGFBR3	IL-10	IL-10	Protein coding
TGFBR3	IL-6	IL-6	Protein coding
TGFBR3	STAT4	STAT4	Protein coding
TGFBR3	TGFB1	Transforming growth factor-β1	Protein coding
HIF1A-AS3	HIF1A	Hypoxia-inducible factor 1 subunit α	Protein coding
HIF1A-AS3	SNAPC1	snRNA-activating complex polypeptide 1	Protein coding
HIF1A-AS3	HIF1A-AS2	HIF1A antisense RNA 2	RNA gene

Run from VarElect (Copyright © LifeMap Sciences, Inc.), used with permission.

HIF1A-AS3: This is an RNA gene and is affiliated with the long noncoding RNA class of molecules. It appears to play an indirect role in the development of RA.[11–13]

Twelve VUSs were identified encoding 11 proteins within the *HIF1A-AS3* gene in patients analyzed. All but one of these variants were classified as missense mutations. One mutation was classified as a synonymous variant.

Discussion

This study provided additional evidence to support the use of incorporating exome sequencing, DNA variant phenotype-driven ranking filters with knowledge engineering via the use of biomimetic digital twins, GeneCards, and VarElect, to provide a greater understanding of the molecular mechanism of disease. Furthermore, these results are beginning to show the value of multiomics and the use of digital twins for enhanced molecular diagnostics and the potential to begin reclassifying VUSs.

The study identified five pathogenic, and one likely pathogenic, DNA variants in 8 of 25 patient samples analyzed, but not in 25 control samples.

Clinical molecular laboratory directors face immense challenges in making decisions on reporting out VUSs. The number of VUSs identified in exome sequencing can vary significantly from person to person. Exome sequencing typically identifies thousands of genetic variants within exons. These variants include single-nucleotide variants, small insertions or deletions, and larger structural variants.

Many of these variants may be common in the population and have been well studied, whereas others may be rare or previously unreported. VUSs are those genetic variants whose significance in relation to disease or health outcomes is not understood. These require further investigation, functional studies, or larger population studies to determine their clinical relevance.

The number of VUSs identified in exome sequencing depends on various factors, including the individual's genetic background, ethnicity, and family history, and the specific criteria used to classify variants as VUSs. Additionally, the depth and accuracy of sequencing, as well as the bioinformatics tools and databases used for variant interpretation, can also influence the number of VUSs identified.

In clinical settings, genetic counselors, geneticists, oncologists, and other health care specialties carefully assess and interpret variants identified through exome sequencing to provide patients with the most accurate information regarding their potential health implications. As our understanding of the human genome and the functional significance of genetic variants continues to evolve, the interpretation of VUSs will also change over time.

Herein, the genotype-phenotype ranking and a biomimetic digital twin engine were used to identify 3172 VUSs potentially associated with the pathophysiology of RA.[11–13]

In addition to the digital twin engine, GeneCards and VarElect were incorporated into this analysis. Four VUSs were identified in genes *HIF1A*, *HLA-DOA*, *PTGER3*, and *HIPK3*, which are directly related to the development of RA[11–13] (Table 2).

The *HIF1A* gene was found in 7 of 25 patient samples, *HLA-DOA* in 4, *PTGER3* VUS in 4, and *HIPK3* VUS in 4.

All but one of the VUSs identified within the *HIF1A* gene were classified as missense mutations. Missense variants are a genetic alteration in which a single base pair substitution alters the genetic code in a way that produces an amino acid that is different from the usual amino acid at that position. Many missense variants will alter the function of the protein and be disease causing.

Hyperplasia of synovial fibroblasts, infiltration with inflammatory cytokines, and tissue hypoxia are major characteristics of RA.[21] IL-33 is an inflammatory cytokine exacerbating the disease severity of RA. HIF-1α (HIF-1A) shows increased expression in RA synovium and could regulate several inflammatory cytokine productions. Elevated levels of IL-33 have been shown in synovial fluids of patients with RA. HIF-1A promotes the activation of the signaling pathways controlling IL-33 production, particularly the p38 and extracellular signal-regulated kinase pathways. IL-33, in turn, could induce more HIF-1α expression, thus forming a HIF-1α/IL-33 regulatory circuit that would perpetuate the inflammatory process in RA.

Three VUSs identified here encoded three different proteins within the *HLA-DOA* gene.

One was a missense mutation, one was a synonymous variant, and one was a start-loss mutation. Start-loss mutations are a point mutation in the ATG start codon of a transcript that reduces or eliminates protein production. The elimination or reduction of a functional protein is most likely a disease-causing DNA variant. A synonymous mutation is a genetic change that alters a gene's DNA sequence but not the protein sequence it encodes. Synonymous mutations have traditionally been considered neutral mutations because they do not change the amino acid that is translated. However, recent studies suggest that synonymous mutations can have a significant impact on RNA stability, RNA folding, translation, and cotranslational protein folding.

Okada et al[22] conducted a large-scale major histocompatibility complex fine-mapping analysis of patients with RA in a Japanese population (6244 RA cases and 23,731 controls) by using HLA imputation, followed by a multi-ethnic validation study including east Asian and European populations ($n = 7097$ and 23,149, respectively). They identified an independent risk of a synonymous mutation at *HLA-DOA*, a non-classic HLA gene, on ACPA-positive RA risk ($P = 1.4 \times 10^{-9}$), which demonstrated a *cis* expression quantitative trait loci effect on *HLA-DOA* expression. Transethnic comparison revealed different linkage disequilibrium patterns in *HLA-DOA* and *HLA-DRB1*, explaining the

observed *HLA-DOA* variant risk heterogeneity among ethnicities, which was most evident in the Japanese population.

Within the *PTGER3* gene, VUSs were transcribed into six different proteins. All these proteins are predicted to be missense mutations. A missense mutation is a DNA change that replaces an amino acid in a protein with a different one. Missense mutations are also known as nonsynonymous mutations. Some missense mutations have little to no effect on the protein's function, whereas others can alter it. For example, a missense mutation in the caveolin-3 gene is associated with limb-girdle muscular dystrophy in humans. Another missense mutation in the *PAX3* gene can cause Klein-Waardenburg syndrome, which includes limb abnormalities. A different missense mutation in the same amino acid residue can cause craniofacial-deafness-hand syndrome, a more severe disorder.

The protein encoded by the *PTGER3* gene is a member of the G-protein−coupled receptor family. This protein is one of four receptors identified for PGE2. This receptor may have many biological functions, which involve digestion, nervous system, kidney reabsorption, and uterine contraction activities. PGE2 is highly expressed in the inflamed joints of RA, and IL-10 and IL-6 are also abundant. PGE2 is a well-known activator of the cAMP signaling pathway, and there is functional cross talk between cAMP signaling and the Janus kinase (Jak)-STAT signaling pathway.[11−13]

Five VUSs that produced five proteins were identified within the *HIPK3* gene. All the encoded proteins were classified as missense mutations.

HIPK3 encodes a homeodomain-interacting protein kinase 3. This enables protein serine/threonine kinase activity. It is involved in mRNA transcription, cell proliferation, inflammation, negative regulation of the apoptotic process, and protein phosphorylation.[11−13]

Overexpression of *HIPK3* protein in immune cells in patients with RA has also been reported.[22]

Two of these VUSs, *TGFBR3* and *HIF1A-AS3*, are indirectly related to the pathophysiology of RA (Table 3).

The *TGFBR3* DNA variant was found in 5 of 25 patient samples, and the *HIF1A-AS3* VUS was also present in 5 of 25 samples.

Twelve VUSs within the *TGFBR3* gene encoded seven different proteins. All these proteins were missense mutations.

This *TGFBR3* locus encodes the TGF-β type III receptor. The encoded receptor is a membrane proteoglycan that often functions as a coreceptor with other TGF-β receptor superfamily members. Ectodomain shedding produces soluble TGFBR3, which may inhibit TGFB signaling. Decreased expression of this receptor has been observed in various cancers. Alternatively spliced transcript variants encoding different isoforms have been identified for this gene. Diseases associated with *TGFBR3* include familial cerebral saccular aneurysm and priapism. Among its related pathways are apoptotic pathways in synovial fibroblasts and negative regulation of fibroblast growth factor receptor 3 signaling.[11−13]

TGFBR3 is indirectly related to the development of RA by interacting with pathways including *TNF*. This gene encodes a multifunctional proinflammatory cytokine that belongs to the tumor necrosis factor (TNF) superfamily.

TGFBR3 also interacts with pathways for IL-6. This gene encodes a cytokine that functions in inflammation and the maturation of B cells. The protein is primarily produced at sites of acute and chronic inflammation, where it is secreted into the serum and induces a transcriptional inflammatory response through IL-6 receptor, α.

TGFBR3 plays an indirect role in the development of RA by interacting with the *TGHB1* gene pathway. This gene encodes a secreted ligand of the TGF-β superfamily of proteins. Ligands of this family bind various TGF-β receptors, leading to recruitment and activation of SMAD family transcription factors that regulate gene expression. The encoded preproprotein is proteolytically processed to generate a latency-associated peptide and a mature peptide and is found in either a latent form composed of a mature peptide homodimer, a latency-associated peptide homodimer, and a latent TGF-β binding protein, or in an active form, consisting solely of the mature peptide homodimer. The mature peptide may also form heterodimers with other TGFB family members. This encoded protein regulates cell proliferation, differentiation, and growth, and can modulate expression and activation of other growth factors.

Gene pathways including *IL-10* play a role in the pathophysiology of RA. *TGHBR3* plays an indirect role in the activation of this pathway. IL-10 encodes a cytokine that is produced primarily by monocytes and to a lesser extent by lymphocytes. This cytokine has pleiotropic effects in immunoregulation and inflammation. It down-regulates the expression of type 1 helper T cell cytokines, major histocompatibility complex class II antigens, and costimulatory molecules on macrophages. It also enhances B-cell survival, proliferation, and antibody production. This cytokine can block NF-κB activity and is involved in the regulation of the JAK-STAT signaling pathway.

TGFBR3 also plays an indirect role in the development of RA by interacting with the Stat-4 pathway. This protein encoded by this gene is a member of the STAT family of transcription factors. In response to cytokines and growth factors, STAT family members are phosphorylated by the receptor-associated kinases, and then form homodimers or heterodimers that translocate to the cell nucleus, where they act as transcription activators. This protein is essential for mediating responses to IL-12 in lymphocytes and regulating the differentiation of T helper cells. DNA variants in this gene may be associated with systemic lupus erythematosus and RA.

Of the 11 encoded proteins from the *HIFA-AS3*, all but one are potentially disease causing and associated with the pathophysiology of RA.

The *HIF1A-AS3* gene is also indirectly related to the pathophysiology of RA by interacting with pathways that include the genes *HIF1A*, *SNAPC1*, and *HIF1A-AS2*.[11−13]

The *HIF1A* gene encodes the α subunit of transcription factor HIF-1, which is a heterodimer composed of an α and a β subunit. HIF-1 functions as a master regulator of cellular and systemic homeostatic response to hypoxia by activating transcription of many genes, including those involved in energy metabolism, angiogenesis, apoptosis, and other genes whose protein products increase oxygen delivery or facilitate metabolic adaptation to hypoxia. HIFs are transcription factors that are activated in response to decreased oxygen availability in the cellular environment. Tissue hypoxia is a major characteristic of RA.

HIF1A-AS3 plays an indirect role in the gene pathway of *SNAPC1*. The *SNAPC1* gene product is a small nuclear RNA activating complex polypeptide 1. It is predicted to enable sequence-specific DNA binding activity. It is also predicted to be involved in snRNA transcription by RNA polymerase II and snRNA transcription by RNA polymerase III. The *SNAPC1* pathway plays a role in the development of RA.

HIF1A-AS2 (HIF1A antisense RNA 2) is an RNA gene and is affiliated with the long noncoding RNA class of molecules. *HIF1A-AS3* potentially plays an indirect role in the pathophysiology of RA by interacting with the *HIF1A-AS2* pathway.

The identification of these VUSs does not confirm that they play a role in the development of RA, but the fact that most of their encoded proteins are classified as dysfunctional strongly suggests that they are highly likely to play some role in the pathophysiology of this disorder.

Proving that a VUS is pathogenic or likely pathogenic involves a comprehensive process of variant interpretation and assessment. This process typically involves multiple steps and considerations, including the following:

i) Clinical: The first step is to gather clinical information about the individual who underwent genetic testing. This includes the individual's medical history, family history, presenting symptoms, and any relevant clinical findings. Understanding the phenotype associated with the variant can provide valuable context for its interpretation.
ii) Classification guidelines: Variants identified through genetic testing are classified according to established guidelines, such as those provided by the American College of Medical Genetics and Genomics or the Association for Molecular Pathology. Variants are categorized into five main classes: pathogenic, likely pathogenic, variants of unknown clinical significance, likely benign, and benign.
iii) Functional studies: These may be conducted to assess the impact of the variant on protein function or expression. These studies can provide direct evidence of the variant's pathogenicity by demonstrating its effect on cellular processes or protein function. *In silico* applications can also be used to determine whether a protein is a functional or a nonfunctional protein.
iv) Population frequency: Variants that are rare in the general population are more likely to be pathogenic, especially if they are found in genes known to be associated with disease. Population databases, such as the Exome Aggregation Consortium or the Genome Aggregation Database, can be used to assess the frequency of the variant in different populations.
v) Segregation: In families with multiple affected individuals, segregation analysis can be used to determine whether the variant cosegregates with the disease phenotype. If the variant is found in all affected family members but not in unaffected individuals, this provides strong evidence of its pathogenicity.
vi) *In silico* predictions: Computational algorithms and bioinformatics tools can be used to predict the functional impact of a variant based on its location within the gene and its effect on protein structure. Although these tools are not definitive proof of pathogenicity, they can provide supporting evidence.

Proving pathogenicity or likely pathogenicity for a VUS is often challenging and may require multiple lines of evidence. In many cases, variants initially classified as VUSs may be reclassified over time as additional evidence becomes available. Therefore, ongoing research and updates to variant databases are essential for improving our understanding of genetic variation and its clinical significance.

The reclassification of VUSs is one of the most significant challenges in genetics today. Recently, there was an industry-sponsored symposium at the American College of Medical Genetics and Genomics 2024 annual conference that discussed the importance of reclassifying VUSs, and how soon this could be achieved. The symposium concluded that it would take approximately 10 to 15 years to accomplish this goal.

One potential limitation of these results is that only 25 patient samples and 25 normal controls were analyzed. However, a biomimetic digital twin analysis is powerful in its ability to analyze small, wide data sets and identify hidden or dark data and unknown biological relationships. Another potential limitation is that 19 of 25 of our patients were White. Additional studies are required to discover the role of ethnicity, if any, in the pathogenesis of RA.

This multiomics and biomimetic digital twins technology is new, and some believe it is controversial and is in competition with AI, ML, and neural networks. We believe that the use of AI, ML, neural networks, and biomimetic digital twin analysis should all be used together to address the multidimensional, multiscale characteristics of problems in modeling and advanced analytics in general, and in biomedicine in particular.

Traditional information technology twin systems use statistical machine learning algorithms on normalized large data sets for classification. AI and ML analyses remove outliers, normalize data, and usually require a training set. The development of the training set could potentially

introduce unintentional bias. The algorithms can find only what they are programmed to look for, and the outcomes are predictions that need to be tested.

Some AI and ML users disagree that they require large data sets for analysis and that they are not a black box. The US Food and Drug Administration, the NAS, the Massachusetts Institute of Technology, and many sources of AI expertise call it a black box because there is no way to audit the statistical computations from token to token that eventually lead to a prediction that must be verified externally. The US Food and Drug Administration also requires contextual interpretation. Contextualization is achieved by modeling multidimensional and multiscale relationships. Computing statistical proximity just cannot do that. It is like autocorrect, it can only predict the next word, and it only gets it right if you keep sending the same messages.

The current omics and biomimetic digital twins research design is not a population-based genetic association case-control study that requires statistical analyses, including odds ratios and *P* values. In this analysis, the data are not cleansed or normalized, the classification is driven by expert knowledge graphs, the software models the complex relationships across diverse small and wide data using the expertise graphs to compute relevance, and the software outputs all the potential relationship scenarios.

In conclusion, the current results suggest that multiomics and biomimetic digital twins can provide more insight into the development of RA. It can also help in the process of reclassifying VUSs potentially associated with the pathophysiology of RA. The reclassification of VUSs will play a critical role in complex molecular diagnostics and drug development.

Disclosure Statement

W.G.K. and L.K. own Genzeva and LumaGene, and class C membership units in RYLTI, LLC, of which RYLTI BioPharma is a subsidiary. L.B. and J.G. own class C membership units in RYLTI, LLC, of which RYLTI BioPharma is a subsidiary. No other author has any financial or personal relationship that could cause a conflict of interest regarding this article.

References

1. Kulkarni PA, Singh H: Artificial intelligence in clinical diagnosis: opportunities, challenges, and hype. JAMA 2023, 330:317−318
2. National Academies of Sciences, Engineering, and Medicine: Opportunities and Challenges for Digital Twins in Biomedical Research. Washington, DC, The National Academies Press, 2022
3. National Academies of Sciences, Engineering, and Medicine: Physics of Life. Washington, DC, The National Academies Press, 2022
4. National Academies of Sciences, Engineering, and Medicine: Foundational Research Gaps and Future Directions for Digital Twins. Washington, DC, The National Academies Press, 2023
5. Kearns WG, Stamoulis G, Glick J, Baisch L, Benner A, Brough D, Du L, Wilson B, Kearns L, NG N, Seshan M, Anchan R: The application of knowledge engineering via the use of a biomimetic digital twin ecosystem, phenotype driven variant analysis, and exome sequencing to understand the molecular mechanisms of disease. J Mol Diagn 2024, 26:543−551
6. Smolen JS, Aletaha D, McInnes IB: Rheumatoid arthritis. Lancet 2016, 388:2023−2038
7. Radu AF, Bungau SG: Management of rheumatoid arthritis: an overview. Cells 2021, 10:2857
8. McInnes IB, Schett G: The pathogenesis of rheumatoid arthritis. N Engl J Med 2011, 365:2205−2219
9. Smith MH, Berman JR: What is rheumatoid arthritis? JAMA 2022, 327:1194
10. Padyukov L: Genetics of rheumatoid arthritis. Semin Immunopathol 2022, 44:47−62
11. Stelzer G, Rosen R, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, Iny Stein T, Nudel R, Lieder I, Mazor Y, Kaplan S, Dahary D, Warshawsky D, Guan - Golan Y, Kohn A, Rappaport N, Safran M, Lancet D: The GeneCards suite: from gene data mining to disease genome sequence analyses. Curr Protoc Bioinformatics 2016, 54:1.30.1−1.30.33
12. Safran M, Rosen N, Twik M, BarShir R, Iny Stein T, Dahary D, Fishilevich S, Lancet D: The GeneCards Suite Practical Guide to Life Science Databases. Singapore, Springer, 2022. pp. 27−56
13. Stelzer G, Plaschkes I, Oz - Levi D, Alkelai A, Olender T, Zimmerman S, Twik M, Belinky F, Fishilevich S, Nudel R, Guan - Golan Y, Warshawsky D, Dahary D, Kohn A, Mazor Y, Kaplan S, Iny Stein T, Baris H, Rappaport N, Safran M, Lancet D: VarElect: the phenotype-based variation prioritizer of the GeneCards suite. BMC Genomics 2016, 17(Suppl 2):444
14. Ni J, Wang P, Yin KJ, Yang XK, Cen H, Sui C, Wu GC, Pan HF: Novel insight into the aetiology of rheumatoid arthritis gained by a cross-tissue transcriptome-wide association study. RMD Open 2022, 8:e002529
15. Goldmann K, Spiliopoulou A, Iakovliev A, Plant D, Nair N, Cubuk C, MATURA Consortium, McKeigue P, Barnes MR, Barton A, Pitzalis C, Lewis MJ: Expression quantitative trait loci analysis in rheumatoid arthritis identifies tissue specific variants associated with severity and outcome. Ann Rheum Dis 2024, 83:288−299
16. National Academies of Sciences, Engineering, and Medicine: Applying Systems Thinking to Regenerative Medicine: Proceedings of a Workshop. Washington, DC, The National Academies Press, 2021
17. National Academies of Sciences, Engineering, and Medicine: Closing Evidence Gaps in Clinical Prevention. Washington, DC, The National Academies Press, 2022
18. Rottman BM, Genter D, Goldwater MB: Causal systems categories: differences in novice and expert categorizations of causal phenomena. Cogn Sci 2012, 36:919−932
19. Spivak DI, Kent RE: Ologs: a categorical framework for knowledge representation. PLoS One 2012, 7:e24274
20. Wasserstein RL, Lazar NA, Lazar NA: The ASA statement on p-values: context, process, and purpose. Am Stat 2016, 70:129−133
21. Hu F, Shi L, Mu R, Zhu J, Li Yingni, Ma X, Li C, Jia R, Yang D, Li Y, Li Z: Hypoxia-inducible factor-1[alpha] and interleukin 33 form a regulatory circuit to perpetuate the inflammation in rheumatoid arthritis. PLoS One 2013, 8:e72650
22. Okada Y, Suzuki A, Ikari K, Terao C, Kochi Y, Ohmura K, Higasa K, Akiyama M, Ashikawa K, Kanai M, Hirata J, Suita N, Teo YY, Xu H, Bae SC, Takahashi A, Momozawa Y, Matsuda K, Momohara S, Taniguchi A, Yamada R, Mimori T, Kubo M, Brown MA, Raychaudhuri S, Matsuda F, Yamanaka H, Kamatani Y, Yamamoto K: Contribution of a non-classical HLA gene, HLA-DOA, to the risk of rheumatoid arthritis. Am J Hum Genet 2016, 99:366−374