JCBI
Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606
Volume 11 Issue 01
2026

Review Article
https://doi.org/10.56979/1101/2026/1514

Knowledge Graphs for Explainable Artificial Intelligence and Decision-Transparency in Genomics: A Review of Methods, Applications, and Challenges

Zainab Azhar[1*], Omer Irshad[1], Abdul Sattar[2], Ali Hussain[3], Muhammad Haseeb Zia[4]

[1]Department of Software Engineering, Lahore Garrison University, Lahore, 54000, Pakistan.
[2]Department of Computer Science, Lahore Garrison University, Lahore, 54000, Pakistan.
[3]Department of Computer Science and Information Technology, The University of Lahore, Lahore, 54000, Pakistan.
[4]Digital Forensic Research and Service Center, Lahore Garrison University, Lahore, 54000, Pakistan.
[*]Corresponding Author: Zainab Azhar. Email: zainabazhar@lgu.edu.pk

Received: February 10, 2026 **Accepted:** May 20, 2026

Abstract: In recent years, the field of Artificial Intelligence has emerged as one of the fastest growing and extensively used technologies in the scientific world, and is transforming the way large scale complex data is processed and understood. In genomics, a field with a ton of different and massive datasets, AI has become an increasingly important tool for speeding up the discovery process, but there are concerns with this and transparency, as well as with the trust AI can bring. To address this need, we performed a structured literature review of the few existing systems, explainability methods, and unaddressed challenges that use genomic knowledge graphs and explainable or comprehensible AI, and compiled a pipeline for the field.The amount of genetic data produced is expanding at a rate far more rapid than can be processed. While this data has great potential to enhance precision medicine, traditional analytical approaches struggle to effectively address the volume and types of sequence variants, gene annotations, multi-omics profiles, and complex regulatory networks. One of the more useful approaches to solve this problem is the use of Knowledge Graphs (KGs), which structure biological information as a network of entities and relationships. These graphs can be given a uniform language by using ontologies.The challenge is that many of the AI models that are created on these knowledge graphs remain as 'black boxes'. They make predictions without giving reasons behind their predictions, which is a major drawback in a discipline such as genomics, where decisions can have direct impact on patient care and must stand up to clinical and regulatory scrutiny. The study explores the efforts being made to bridge that gap. We discuss the most prominent KG systems currently available, explore strategies under development to increase the interpretability of AI predictions and assess the remaining challenges.

Keywords: Comprehensible Artificial Intelligence; Black-Box Models; Genomic Data Integration; Knowledge Representation

1. Introduction

Artificial Intelligence (AI) has become a transformative and one of the most prevalent and rapidly changing technologies in scientific research today. With its pattern recognition, modeling of complex relationships, and predictive capabilities from vast amounts of data, it has been widely used in nearly all data-rich disciplines, including healthcare, finance, climate science, and materials discovery. This growth is largely due both to improvement in model architectures (e.g., deep learning, graph-based learning) and to the proliferation of large, structured and semantically rich datasets on which models can be trained. This change has had a

significant impact on one of the areas most transformed by genetic engineering genomics [1]. There is no other field that can match the scale and complexity of the field of genetics and the insights that can be gained from the data, and researchers have been increasingly relying on AI to speed up various aspects of research, including the discovery of new variants and the identification of genes responsible for the emergence of diseases [2]. Alongside this same quick adoption has also emerged a parallel but equally urgent question: how do increasingly powerful AI models predict outcomes, and why? The models are becoming more opaque, and more and more often they fail to provide enough information to explain their conclusions. This opacity is not acceptable in a field such as genomics, where the results could directly impact clinical practice and the health of patients, and where trust, adoption and regulatory approval will be essential. This survey exists because of this intersection between the increasing influence of AI and the increasing demand for transparency. To fill this need, we review the related works systematically to summarize systems, explainability techniques, and challenges in the space of genomic knowledge graphs and explainable/comprehensible AI in a coherent roadmap for the field.

Despite the rapid expansion of genetic data has opened up a lot of possibilities for improving precision medicine and biological research, handling and analyzing such intricate, multi-omics datasets is still quite difficult [3]. Large volumes of data are produced by high-throughput sequencing, spanning from regulatory networks to gene annotations, requiring transparent and organized methods for integration and analysis. By organizing biological knowledge into nodes (entities) and labeled edges (relationships), Knowledge Graphs (KGs) improve semantic consistency, reasoning, and interoperability. The language and interactions between genes, variations, pathways [4], and diseases are standardized by foundational ontologies as the Gene Ontology (GO), Genotype Ontology (GENO), and Ontology of Genes and Genomes (OGG) [5,6].The integrative potential of such frameworks in connecting data to useful information is demonstrated by genomic KGs such as DisGeNET [7,8], Hetionet [9], KnetMiner [10], and VariantKG [11].Although most AI models that use these KGs function as "black boxes," which reduces regulatory acceptability and trust in clinical genomics. For dependable, data-supported, and clinically interpretable outcomes, it is crucial to guarantee transparency and explainability in AI-driven predictions [12, 13].

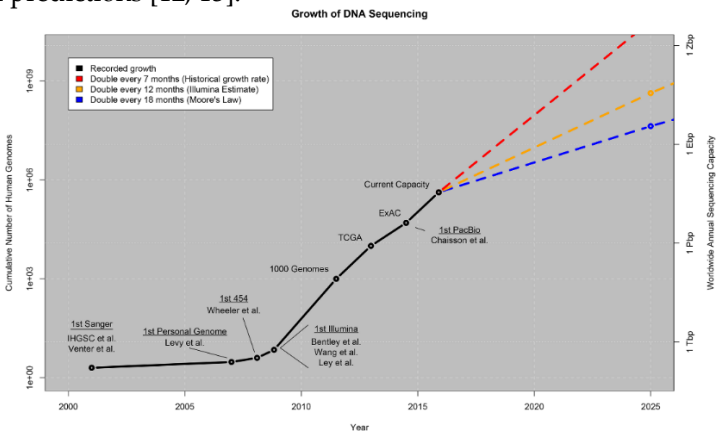

Figure 1. Growth of DNA sequencing

The development of DNA sequencing hasn't been linear as you can see in Figure 1. The first Sanger-sequenced human genome (International Human Genome Sequencing Consortium, 2001) and the first personal genome both took a number of years to accomplish, but the advent of next-generation sequencing platforms led to an exponential increase in output. By 2015, cumulative sequencing capacity had increased from a few hundred genomes in 2001 to several hundred thousand genomes, with large-scale initiatives like the 1000 Genomes Project, The Cancer Genome Atlas (TCGA) and the Exome Aggregation Consortium (ExAC) driving this growth. Based on three reasonable doubling time projections (7 months, 12 months, and 18 months), predicted that by the mid-2020s, global sequencing capacity will be anywhere from one exabase to one zettabase per year. This path highlights the need for a solution to the problem this survey aims to tackle: scalable and interpretable AI systems that can be developed on structured knowledge representations like knowledge

graphs, or else we will be overwhelmed by genomic data and unable to garner meaningful, trustworthy biological knowledge from them [14].

The purpose of Explainable and Comprehensible AI (XAI/CAI) is to make reasoning processes comprehensible. According to KGs offer a semantically rich foundation for explanations that are provenance-enabled and context-aware. Techniques like rule-based reasoning [15], graph-based subgraph or path extraction [16], model-agnostic techniques [17], and hybrid neuro-symbolic integration [18] have demonstrated promise in improving interpretability without compromising accuracy. Trust in high-stakes fields like cancer genomics is further reinforced by provenance-enabled KGs [15]. Despite these developments, there are still difficulties, especially when it comes to modeling intricate biological processes [19], integrating heterogeneous, noisy data [20], and striking a balance between predictability and transparency [21, 22]. Future research must focus on developing scalable, explainable AI systems that can reason efficiently over massive genomic KGs [23].

In order to tackle these issues, this survey summarizes recent findings at the nexus of transparent AI decision-making and genetic knowledge graphs. In particular, we (1) examine current genomic KGs and systems that operationalize them, (2) examine frameworks and techniques that use KGs to facilitate CAI, such as provenance tracking, rule-mining, path-finding, and embedding-based explainability, and (3) identify new research areas and promising techniques for developing reliable, interpretable, and efficient genomic AI systems [15, 24]. The goal of this survey is to present a roadmap for developing transparent genomic AI that is both therapeutically significant and scientifically sound by combining lessons from ontology design, KG system development, and AI transparency.

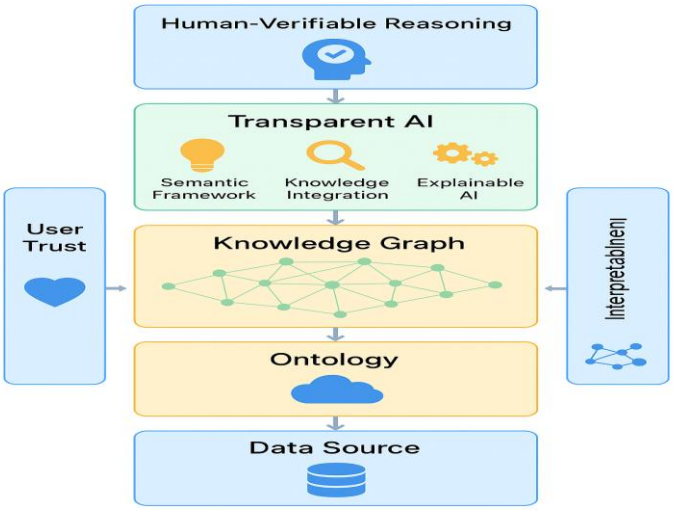

Figure 2. Flow of Transparent AI

In the figure 2. illustrates the flow of structured knowledge from Ontology to Knowledge Graph to Transparent AI, resulting in an AI system that is reliable and comprehensible. It demonstrates how Explainable AI, Knowledge Integration, and Semantic Framework combine to convert unstructured data into clear, insightful understandings.

2. Genomic Ontologies, Building Blocks of Knowledge Graphs

A semantically rich genomics Knowledge Graph (KG) requires standardized ontologies. Genes, variations, pathways [4], cohorts, and genome editing methods are all represented by consistent vocabularies that they give. These ontologies serve as the foundation of the KG, guaranteeing accurate integration of various types of data.

2.1. Gene and Gene Product Ontologies

These ontologies represent genes and their functions. Gene Ontology (GO) describes Biological Process, Molecular Function, and Cellular Component [5, 25] Ontology of Genes and Genomes (OGG) and its mouse

subset OGG-MM define genes, families, and genomes across species. HUGO Gene Nomenclature (HGNC) standardizes human gene symbols and identifiers [26]. Genotype Ontology (GENO) models genotypes, alleles, and their links to phenotypes and diseases.

2.2. Sequence Variation and Polymorphism Ontologies

This group analyzes the impacts of genetic variation. The Genome Variation Ontology (GVO) encompasses structural variants, CNVs, indels, and SNPs. SNPO, or Single-Nucleotide Polymorphism Ontology, is centered on SNP characteristics [27]. By linking them to genes and regulatory elements, the Ontology for Genetic Interval (OGI) and Genomic Feature [28] and Variation Ontology (GFVO) model variants and genomic areas.

2.3. Genome Features and Components

The properties and components of the genome are described by these ontologies. Chromosomes, plasmids, and organellar genomes are defined by the Genome Component Ontology (GCO). Genes, transcripts, proteins, regulatory elements, and experimental annotations are represented by the Genome Biology Ontology Language (GBOL[29]) and Genetic Feature Format Ontology (GFFO).

2.4. Pathway and Biological Process Ontologies

The pathways that connect genes, proteins, enzymes, and metabolites to their corresponding biological activities are represented hierarchically by the Pathway Ontology [4]

2.5. Ontologies of Population, Cohort, and Ancestry

Human populations and study cohorts are described by ontologies in this category. Genomics Cohorts Knowledge Ontology models bio sample metadata [30], research design, demographics, and cohort size. Population terminology and ancestry are standardized by the Human Ancestry Ontology (HANCESTRO) [31].

2.6. Ontologies of Microbes and Resistance

Resistance genes, mutations, mechanisms, and their effects on antibiotics are all modeled by the Antibiotic Resistance Ontology (ARO) [32, 33].

2.7. Ontologies for Metadata and Genome Editing

This team records FAIR data standards and genome editing procedures. Tools, procedures, and results of genome editing are defined by the NIST Genome Editing Lexicon (NIST_GEL). In accordance with FAIR principles, FAIR Genomes (FG) guarantees structured information for sequencing data [34].

Table 1. Building Blocks of Knowledge Graphs

Knowledge Graph (Ontology)	**Knowledge Graph (Ontology)**
Description	**Description**
Gene Ontology (GO)	An ontology for describing the function of genes and gene products.
Genotype Ontology (GENO)	An integrated ontology for representing the genetic variations described in genotypes, and their causal relationships to phenotype and diseases.
Ontology of Genes and Genomes (OGG)	A formal ontology of genes and genomes of biological organisms.
Genomics Cohorts Knowledge Ontology	An ontology to represent genomics cohort attributes.
Antibiotic Resistance Ontology	Represents antibiotic resistance genes and mutations.
Pathway Ontology	A controlled vocabulary for annotating gene products to pathways.

FAIR Genomes (FG)	A semantic metadata schema to power reuse of NGS data in research and healthcare [34].
Genome Variation Ontology (GVO)	Ontology for systematic description of various genomic variations, including complex structural variations in genomes.
Ontology of Genes and Genomes – Mouse (OGG-MM)	Subset of OGG specific to Mus musculus (mouse).
Genome Component Ontology (GCO)	Defines abstract divisions of the total genetic information of an organism by physical separation into components.
NIST Genome Editing Lexicon (NIST_GEL)	Ontology for genome editing technology concepts, supporting biotechnology applications.
GBOL (Genome Biology Ontology Language)	Ontology framework for genome biology [29].
Genetic Feature Format Ontology (GFFO)	Represents genome feature annotations in GFF3 format.
HUGO Gene Nomenclature (HGNC)	Standardized gene names and symbols approved by the HUGO Gene Nomenclature Committee [26].
Human Ancestry Ontology (HANCESTRO)	Provides a systematic description of ancestry concepts used in GWAS studies [31].
Genomic Feature and Variation Ontology (GFVO)	Represents genomic data using RDF or Linked Data formats such as JSON-LD [28].
Single-Nucleotide Polymorphism Ontology (SNPO)	A domain ontology providing a formal OWL-DL representation of genomic variations (SNPs) [27].
Ontology for Genetic Interval (OGI)	Formalizes genomic elements by defining a class 'genetic interval', based on BFO.

3. Systems/Platforms Built on Genomic KGs

Biomedical Knowledge Graph Systems, Purpose and Platforms, these biomedical Knowledge Graphs (KGs) are powered by ontology-based frameworks that combine various genetic, phenotypic, and drug-related data into easily accessible, interpretable, and queryable resources. These KGs aid in the investigation of biological knowledge, drug repurposing, variation interpretation, and gene-disease finding.

3.1. Gene–Disease Discovery

Systems devoted to discovering connections between genes and illnesses

- DisGeNET: connects genes, variations, and diseases by combining literature, experimental data, and carefully chosen sources [7].
- BOCK: creates a gene-centric KG by combining the networks of OLIDA, GO, HPO, co-expression, and PPI.
- ARBOCK: allows BOCK to be extended for mining meta path rules that link gene pairs in biological networks [30].
- Decision Set (DS) Models: BOCK/ARBOCK rule-based interpretable models for predicting hazardous gene pairings [35].
- DiGePred: uses phenotypes and gene characteristics to predict gene pairings that cause digenic diseases [36].
- BioGraph / BioKB: Heterogeneous KG for the finding of gene-disease associations on a wide scale [7, 23].

3.2. Pharmacogenomics and Drug Repurposing

Platforms for pharmacogenomic research, pathway-based reasoning, and drug-target analysis.

- Hetionet: incorporates phenotypes, routes, medications, diseases, genes, and negative impacts.
- RPath: Uses gene-disease pathways to prioritize therapeutic candidates (based on Hetionet [37].
- PharmGKB: Provides pharmacogenomic insights by mapping genes, variations, medications, and illnesses [38].
- OpenBioLink: A benchmark KG for biomedical network link prediction testing [39].

3.3. Prediction at the Variant Level

KGs that highlight genetic variations and the effects they have on phenotype.

- ClinVar KG, VariantKG, GenCC, and PKG: Support pathogenicity predictions, represent variations, and associate them with phenotypes.

3.4. Exploration and Visualization Tools

Platforms for genomic KG visualization, analysis, and querying.

- KnetMiner: FAIR-compliant tool integrating genome-scale data for gene-trait analysis. KnetMiner: a comprehensive approach for supporting evidence-based gene discovery and complex trait analysis across speciesKnetMiner: a comprehensive approach for supporting evidence-based gene discovery and complex trait analysis across species
- Genephony: Local KG that conducts exploratory analysis on genes, SNPs, and transcripts [40].
- GenoLink: A graph-based application for interactively visualizing phenotypes, pathways, genes, and variations [41].

3.5. Knowledge Graph Embedding / ML Tools

Models and tools for KG predictive analysis and embedding learning.

- Metapath2Vec [30], BioKG2vec [42], and Dlemb [43]: embedding methods that capture graph structure and semantic routes.
- TransE and RotatE.: Gene-disease, drug-target, and pathway interactions can be predicted using Vector-space KG models [44, 45].
- R-GCN: A graph neural network model for biological KGs that manages diverse relations

Table 2. Systems Built on Genomic KGs

System / Platform Based on / Description Purpose / Function	System / Platform Based on / Description Purpose / Function	System / Platform Based on / Description Purpose / Function
BOCK (Biological networks and Oligogenic Combinations as a KG)	Integrates OLIDA with gene-centric networks (PPI, co-expression, GO, HPO)	Builds genome-focused KG to study disease-causing gene interactions
ARBOCK	Built on BOCK KG	Mines metapath rules to predict pathogenic gene interactions
Decision Set (DS) Models	Trained from ARBOCK + BOCK KG	Predicts if gene pairs are pathogenic with interpretable rules
DiGePred	Uses gene & phenotype features (some KG-derived)	Predicts digenic gene interactions
Hetionet	Heterogeneous biomedical KG (genes, drugs, diseases, pathways)	Disease–gene prediction, drug repurposing
RPath	Uses Hetionet KG	KG reasoning for drug prioritization via gene–disease paths

BioGraph	Heterogeneous biomedical KG	Genome-scale discovery and disease–gene prediction
BioKB	Combines text-mined and curated biomedical knowledge	Gene interactions and pathway knowledge exploration.
DisGeNET	Integrates gene–disease associations and variants from curated and experimental sources	Understanding gene–disease relationships
PharmGKB	Pharmacogenomics KG (genes, variants, drugs, diseases)	Genome-level gene–drug–disease interactions
OpenBioLink	Benchmark biomedical KG with gene-level entities	Evaluation of link prediction tasks
Metapath2Vec	Embedding learning for heterogeneous biomedical KGs	Predicting gene–disease associations
BioKG2vec	Random-walk-based KG embedding algorithm	Predicting gene–disease associations
Dlemb	Shallow neural network for KG entity embeddings	Predicting gene–disease associations
RotatE	KG embedding model (complex space)	Link prediction tasks including gene–disease
TransE	Translational embedding model for KGs	Gene–disease association prediction
R-GCN	Relational Graph Convolutional Network	Predicting gene–disease associations in KGs
Genephony	Local KG of genomic entities (genes, SNPs, transcripts)	Manipulating and exploring genome entity relationships
GenoLink	Graph querying and visualization tool for genomic data	Functional genomics data exploration
KnetMiner	FAIR KG integrating genome-scale and cross-species data	Gene discovery and trait analysis

4. Transparency Techniques for Genome Knowledge Graphs

Transparency in AI-driven genomic KGs is essential for clinical decision-making and scientific credibility. Techniques that uncover the reasons behind predictions and connect them to interpretable biomedical information can be broadly divided into four categories: model-agnostic explainability, graph-based explainability, biological validation, and hybrid reasoning.

4.1. Model-Agnostic Explainability

Methods for emphasizing key properties that apply to any AI model.

- By approximating complex models with interpretable ones, LIME explains predictions locally and is used to find significant genes, variations, or pathways.
- In order to improve clinical interpretability, feature attribution quantifies the input characteristics' (such as gene connections, variants, and illness linkages) contribution to a prediction [46].

4.2. Graph-Based Explainability

Methods for making forecasts interpretable by utilizing the KG structure.

- Subgraph/Path Extraction: Indicates pertinent KG segments that connect genes, proteins, or pathways to anticipated results [47].
- Rule Extraction: Produces logical rules that are readable by humans (e.g., gene X interacts with Y → disease Z) [48, 49]).

- Provenance: Maintains track of the data sources and supporting documentation for every prediction, guaranteeing accountability [50].
- Biological Validation
  Techniques that compare known biological knowledge with AI predictions.
- Pathway Enrichment Validation: increase confidence in results by confirming that predicted genes or variations match established biological pathways [51].

4.3. Hybrid Reasoning

Enables interpretable inference by fusing AI models with symbolic logic.

- Symbolic Reasoning :Generates explicable inference paths by applying KG's logical framework.
- Neuro-Symbolic Integration: Combines symbolic thinking and deep learning to preserve prediction ability while preserving intelligible explanations.

Table 3. Techniques for Genome Knowledge Graphs

Paper	Year	Gnome KG Used	Techniques Used	Explaination of Techniques
Ribeiro et al., "Why Should I Trust You?"	2016	Not applicable (focuses on general AI models)	LIME(Local Interpretable Model-agnostic Explanations	LIME:A method that explains individual predictions by approximating the model locally with an interpretable one.
Murakami et al., Pathogenicity Prediction of Gene Fusions	2024	Fusion-gene KG (fusion genes, pathways, protein domains, miRNAs, references)	LIME, Natural Language Explanations (LLM-generated), KG-based evidence highlighting	LIME: Shows which features of the input influenced the AI's prediction most. Natural Language Explanation: Converts KG evidence into a short, human-readable sentence. KG Evidence Highlighting: Points out the specific paths/facts in the KG that support the prediction.
Abe et al., Explainable AI for Variant Pathogenicity	2023	Genomic/clinical KG (variants, genes, diseases)	Feature importance / attribution	Highlights which KG-derived features (gene connections, variant info, disease links) contributed most to the AI prediction [52].
Renaux et al., Predicting Disease-Causing Genes	2023	Biomedical KG (gene interactions, pathways)	Subgraph / Path extraction	Extracts a small network (subgraph) showing the chain of connections from a gene to disease. Makes it easy to see the reasoning behind the prediction.
Chatzianastasis et al., EMGNN for Cancer Gene Prediction	2023	Biological networks (treated as KG)	Attention-based feature importance, Pathway enrichment validation	Attention: Shows which nodes/genes in the network influenced the prediction most. Pathway enrichment validation: Checks whether predicted genes are part of known biological pathways to make results interpretable.

GenomicKB: a knowledge graph for the human genome	2023	GenomicKB (all human genes, variants, regulatory info)	Provenance / Evidence linking	Shows exactly which database, paper, or source supports a prediction. Helps users trace back the AI decision to real, trustworthy evidence.
Chandak et al., Building a KG for Precision Medicine	2023	Large integrative biomedical KG	Provenance / Subgraph evidence	Combines showing the path in the KG used for the prediction with the source of the facts, so the reasoning behind the AI's output is transparent.
Rajabi et al., Knowledge-graph-based explainable AI: A systematic review	2022	Various biomedical KGs	Rule extraction, Subgraph extraction, Pathway enrichment	Rule extraction: Derives logical rules from the KG to explain decisions. Subgraph extraction: Identifies relevant subgraphs in the KG that support predictions. Pathway enrichment: Assesses whether predicted genes are involved in known biological pathways.
Lecue et al., On the Role of Knowledge Graphs in Explainable AI	2019	Various KGs	Rule-based reasoning, Subgraph extraction	Rule-based reasoning: Applies logical rules derived from the KG to explain AI decisions. Subgraph extraction: Identifies and uses relevant subgraphs in the KG to support predictions.

5. Integration: Ontology → KG System → Transparent AI

Transparency approaches are the last step in a multi-layered pipeline that starts with ontologies, operationalizes through KG systems, and integrates transparent AI into genomic applications. Ontology design and KG construction have a direct impact on the quality and interpretability of the final AI outputs because each layer builds upon the one before it.

5.1. Ontology Layer:

Systems for genome-based KG, such as DisGeNET, which combines experimental and curated gene-disease correlations Hetionet, a drug repurposing platform that integrates medications, genes, and diseases into a diverse biological network. KnetMiner makes it possible to generate hypotheses for complicated trait genomics, In addition to the more current VariantKG, which offers scalable genomic variant representation for graph learning.

5.2. Transparent AI Layer:

Transparency approaches, such as LIME, which locally approximates complex models to generate interpretable forecasts, are used to AI predictions at the top of the pipeline subgraph and path extraction, which finds relational chains of biological significance to support outputs [53, 54], provenance tracking, which traces predictions back to data sources for traceability, and rule-based reasoning, which supplements embedding-based techniques with symbolic explanation [55-58],These methods are particularly important in clinical genomics, where predictability and interpretability are improved by transparency.

This layered perspective highlights the fact that transparency only becomes apparent at the very end and is inevitably reliant on the caliber of the KG system building and ontology design. The efficacy of downstream explainability techniques is directly hampered by errors or inconsistencies in earlier phases, highlighting the significance of an integrated, end-to-end design for transparent genomic AI.

6. Research Gaps & Open Challenges

Even with advancements in the use of Knowledge Graphs (KGs) for transparent AI in genomics, several significant obstacles still need to be overcome. The first is the absence of defined explainability frameworks in genomic AI. The lack of a unified paradigm for creating or assessing explainability in this field leads to fragmented, non-reproducible methods and emphasizes the need for domain-specific standards, even though numerous studies suggest methods like feature attribution, path extraction, and provenance tracking [59]. Second, most existing transparency strategies are still correlation-based, concentrating on identifying significant genes or variations without revealing the underlying biological or causal mechanisms. New frameworks for causal inference for omics data show promise for progressing from correlation to mechanistic interpretability, a field in genomic AI that is still largely uncharted [60-62].

Third, clinical adoption necessitates provenance, validation, and confidence in addition to accuracy. Clinicians' trust and interpretability are increased when model predictions are connected to carefully chosen, physiologically validated sources (such as variant–gene–disease linkages in KGs. Lastly, there is still a lack of research on the integration of neuro-symbolic thinking with genomic KGs. Although explainability has been enhanced in other domains by these hybrid approaches, which combine neural embeddings with symbolic ontologies and rules, their use in clinical genomics, disease mechanism discovery, and variant pathogenicity prediction is still in its infancy [63].

7. Analysis and Synthesis

The ontologies, systems, and explainability techniques surveyed in this paper together suggest a space that is composed of many components but not structurally connected between the layers. At the ontology level, the eighteen resources reviewed cover a wide range from general foundational vocabularies, like GO, GENO, and OGG, to more closely-tuned ones, such as ANTIBIOTIC RESISTANCE, NIST_GEL for genome editing, and HANCESTRO for human ancestry. This diversity reflects the high investment in semantic standardization within the community but also shows that there is no consolidation toward a smaller number of interoperable standards in genomic knowledge. A KG constructed with the sole purpose of storing GO does not necessarily have the granularity level of GVO, and a KG constructed with the sole purpose of storing HANCESTRO does not necessarily have the context of the ancestors of HANCESTRO. Integrating multiple ontologies is still a manual, project-specific process and is not a solved problem.

A similar fragmentation can be seen at the systems level, but along a different dimension. The platforms reviewed can be divided into three functional categories: rule-based, interpretable-by-design, (ARBOCK's metapath rules, Decision Set models), large curated discovery platforms that focus on coverage rather than transparency (DisGeNET, Hetionet, PharmGKB, KnetMiner), and embedding-based models that are inherently opaque by construction (TransE, RotatE, R-GCN, Metapath2Vec, BioKG2vec, Dlemb). Importantly, most of the systems reviewed are in the bottom two categories, as most of the currently operating genomic KG systems are currently not engineered for explainability as the primary architectural goal but rather for maximize coverage of prediction or embedding accuracy.This is where it is evident that there is a real difference. When comparing the explainability techniques they reviewed with the systems they reviewed, few of the methods like LIME, subgraph extraction, provenance tracking, and rule extraction are applied to any of the well-established systems.

None of the techniques explored in this paper are illustrated in a general paper, but instead are shown in a stand-alone paper on a purpose-built or task-specific KG for which the paper was written. No major, widely used, systems were described anywhere in the literature reviewed that had an explainability technique as an inbuilt feature. To sum up, while the field has independently developed its explainability toolkit and

production-grade KG platforms, they have yet to be combined: explainability is not something that users of the big platforms can access directly.A partial instructive exception is the ability to trace back to the origin of an item, which is known as provenance tracking. In GenomicKB and also in the precision medicine KG developed, evidence traceability is integrated during the design phase of the KG and is not "added on" later. This demonstrates that in practical terms, there is no issue of methods; indeed there is no technical reason why provenance-by-design is not possible in other systems, it is just a matter of priorities.A parallel imbalance is seen between the explainability techniques.

As for the current literature, subgraph and rule extraction are the most prevalent techniques, whereas the use of neuro-symbolic and symbolic reasoning techniques, is relatively uncommon. This is typical of one of the key gaps presented in the previous part of this survey. In principle, neuro-symbolic reasoning is particularly appropriate for genomic KGs as it allows the combination of the structured logic provided by ontological rules with the pattern recognition abilities of embedding-based learning, but is also the least explored explainability category among the reviewed works.Together these observations suggest a single takeaway point: the lack of ontologies, systems, or explainability techniques isn't the problem with genomic AI; there is a lack of integration between the three layers of each. Explainability research and production-grade KG systems have taken almost parallel and non-intersecting paths. The goal of closing the gap is probably the most important task for the field in the future, rather than finding new individual techniques.

8. Conclusion

A revolutionary approach to furthering genomic research and therapeutic applications is the combination of explainable AI with Knowledge Graphs (KGs). As the analysis demonstrates, while there are many ontologies, systems and explainability techniques available, they are largely not integrated into the production-grade KG platforms that researchers and clinicians actually use. While explainable AI techniques guarantee that the insights obtained from complicated biological information are transparent, reliable, and therapeutically useful, KGs offer the structural framework for organizing such knowledge. This collaboration is especially important in the field of genomics, where choices have the potential to directly affect patient outcomes and necessitate not only predicted accuracy but also interpretability and validation.

In order to move forward, the science needs to move beyond opaque "black-box" models to systems that represent biomedical reasoning that can be verified by humans. Making sure that outputs are supported by causal, physiologically based explanations necessitates integrating transparency into each step of the pipeline, from ontology design and KG generation to AI-driven inference.

In addition to improving clinical confidence and regulatory acceptance, these advancements will provide researchers and practitioners with the means to connect unprocessed genomic data with useful medical insights.

Future genomics systems can go beyond performance metrics alone by combining rigorous KG frameworks with explainable and interpretable AI to provide solutions that are visible, auditable, and consistent with the fundamental ideas of biomedical research.

References

1. Pakpahan I, Sihombing M, Liu H, Wang M, Su Z, Fang M. Harnessing Artificial Intelligence for Genomic Variant Prediction: Advances, Challenges, and Future Directions. Gigascience. 2026
2. Prince S, Savulescu J. When is black-box AI justifiable to use in healthcare? Big Data Soc. 2025;12(4).
3. Perdomo-Quinteiro P, Belmonte-Hernández A. Knowledge graphs for drug repurposing: a review of databases and methods. Brief Bioinform. 2024;25(6):bbae461. doi:10.1093/bib/bbae461.
4. Petri V, Jayaraman P, Tutaj M, Hayman GT, Smith JR, De Pons J, et al. The pathway ontology - updates and applications. J Biomed Semantics. 2014;5(1).
5. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: Tool for the unification of biology. Nat Genet. 2000;25(1):25–9.
6. Schriml LM, Munro JB, Schor M, et al. The Human Disease Ontology 2022 update. Nucleic Acids Res. 2022;50(D1):D1255-D1261.
7. Piñero J, Bravo Á, Queralt-Rosinach N, Gutiérrez-Sacristán A, Deu-Pons J, Centeno E, et al. DisGeNET: A comprehensive platform integrating information on human disease-associated genes and variants. Nucleic Acids Res. 2017;45(D1):D833–9.
8. Sheils TK, Mathias SL, Kelleher KJ, et al. TCRD and Pharos 2021: mining the human proteome for disease biology. Nucleic Acids Res. 2021;49(D1):D1334-D1346.
9. Himmelstein DS, Lizee A, Hessler C, Brueggeman L, Chen SL, Hadley D, et al. Systematic integration of biomedical knowledge prioritizes drugs for repurposing. Elife. 2017;6.
10. Hassani-Pak K, Singh A, Brandizi M, Hearnshaw J, Parsons JD, Amberkar S, et al. KnetMiner: a comprehensive approach for supporting evidence-based gene discovery and complex trait analysis across species. Plant Biotechnol J. 2021;19(8):1670–8.
11. Prasanna S, Kumar A, Rao D, Simoes EJ, Rao P. A scalable tool for analyzing genomic variants of humans using knowledge graphs and graph machine learning. Front Big Data. 2024;7.
12. Allen B. The Promise of Explainable AI in Digital Health for Precision Medicine: A Systematic Review. J Pers Med. 2024;14(3).
13. Rosenbacke R, Melhus Å, McKee M, Stuckler D. How Explainable Artificial Intelligence Can Increase or Decrease Clinicians' Trust in AI Applications in Health Care: Systematic Review. Jmir Ai. 2024;3.
14. Stephens ZD, Lee SY, Faghri F, Campbell RH, Zhai C, Efron MJ, et al. Big data: Astronomical or genomical? PLoS Biol. 2015;13(7).
15. Schramm S, Wehner C, Schmid U. Comprehensible Artificial Intelligence on Knowledge Graphs: A survey. J Web Semant. 2023;79.
16. Yang H, He G, Zhang M, et al. OntoTiger: a platform of ontology-based application tools for integrative biomedical exploration. Nucleic Acids Res. 2025;1(11). doi:10.1093/nar/gkaf337.
17. Ribeiro MT, Singh S, Guestrin C. "Why should i trust you?" Explaining the predictions of any classifier. Proc ACM SIGKDD Int Conf Knowl Discov Data Min. 2016;13-17-Augu:1135–44.
18. Irshad O, Ghani U. Formalization and Semantic Integration of Heterogeneous Omics Annotations for Exploratory Searches. Current Bioinformatics 2020; 15: pp. 1-17. DOI: 10.2174/1574893615666200127122818.
19. Cao ZJ, Gao G. Multi-omics single-cell data integration and regulatory inference with graph-linked embedding. Nat Biotechnol. 2022;40(10):1458–66.
20. Callahan TJ, Tripodi IJ, Stefanski AL, Cappelletti L, Taneja SB, Wyrwa JM, et al. An open source knowledge graph ecosystem for the life sciences. Sci Data. 2024;11(1).
21. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell. 2019;1(5):206–15.
22. DeLong LN, Mir RF, Fleuriot JD. Neurosymbolic AI for Reasoning Over Knowledge Graphs: A Survey. IEEE Trans Neural Networks Learn Syst. 2025;36(5):7822–42.
23. Lin Y, He J, Chen J, Zhu X, Zheng J, Bo T. BioGraphFusion: graph knowledge embedding for biological completion and reasoning. Bioinformatics. 2025;41(7).
24. Hogan A, Blomqvist E, Cochez M, D'Amato C, Melo G De, Gutierrez C, et al. Knowledge graphs. ACM Comput Surv.

2022;54(4).
25. Aleksander SA, Balhoff J, Carbon S, Cherry JM, Drabkin HJ, Ebert D, et al. The Gene Ontology knowledgebase in 2023. Genetics. 2023;224(1).
26. Seal RL, Braschi B, Gray K, Jones TEM, Tweedie S, Haim-Vilmovsky L, et al. Genenames.org: the HGNC resources in 2023. Nucleic Acids Res. 2023;51(D1):D1003–9.
27. Coulet A, Smaïl-Tabbone M, Benlian P, Napoli A, Devignes MD. Ontology-guided data preparation for discovering genotype-phenotype relationships. BMC Bioinformatics. 2008;9(SUPPL. 4).
28. Baran J, Durgahee BSB, Eilbeck K, Antezana E, Hoehndorf R, Dumontier M. GFVO: The Genomic Feature and Variation Ontology. PeerJ. 2015;2015(5).
29. van Dam JCJ, Koehorst JJ, Vik JO, Martins dos Santos VAP, Schaap PJ, Suarez-Diez M. The Empusa code generator and its application to GBOL, an extendable ontology for genome annotation. Sci Data. 2019;6(1).
30. Irshad O, Ghani U. Integration and Querying of Heterogeneous Omics Semantic Annotations for Biomedical and Biomolecular Knowledge Discovery. Current Bioinformatics 2019; 15(1): pp. 41-58. DOI: 10.2174/1574893614666190409112025.
31. Morales J, Welter D, Bowler EH, Cerezo M, Harris LW, McMahon AC, et al. A standardized framework for representation of ancestry data in genomics studies, with application to the NHGRI-EBI GWAS Catalog. Genome Biol. 2018;19(1).
32. Alcock BP, Huynh W, Chalil R, Smith KW, Raphenya AR, Wlodarski MA, et al. CARD 2023: expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database. Nucleic Acids Res. 2023;51(1 D):D690–9.
33. McArthur AG, Waglechner N, Nizam F, Yan A, Azad MA, Baylay AJ, et al. The comprehensive antibiotic resistance database. Antimicrob Agents Chemother. 2013;57(7):3348–57.
34. van der Velde KJ, Singh G, Kaliyaperumal R, Liao XF, de Ridder S, Rebers S, et al. FAIR Genomes metadata schema promoting Next Generation Sequencing data reuse in Dutch healthcare and research. Sci Data. 2022;9(1).
35. Renaux A, Terwagne C, Cochez M, Tiddi I, Nowé A, Lenaerts T. A knowledge graph approach to predict and interpret disease-causing gene interactions. BMC Bioinformatics [Internet]. 2023;24(1). Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC10463539/?utm_source=chatgpt.com
36. Turner TN, Wilfert AB, Bakken TE, Bernier RA, Pepper MR, Zhang Z, et al. Sex-Based Analysis of De Novo Variants in Neurodevelopmental Disorders. Am J Hum Genet. 2019;105(6):1274–85.
37. Dotolo S, Marabotti A, Facchiano A, Tagliaferri R. A review on drug repurposing applicable to COVID-19. Brief Bioinform. 2021;22(2):726–41.
38. Huang Z, Shi J, Gao Y, Cui C, Zhang S, Li J, et al. HMDD v3.0: a database for experimentally supported human microRNA disease associations. Nucleic Acids Res. 2018;47(D1):D1013-D1017.
39. Breit A, Ott S, Agibetov A, Samwald M. OpenBioLink: A benchmarking framework for large-scale biomedical link prediction. Bioinformatics. 2020;36(13):4097–8.
40. Ben-Yaacov E, Eldar YC. A fast and flexible method for the segmentation of aCGH data. Bioinformatics. 2008;24(16).
41. Durand P, Labarre L, Meil A, Divol JL, Vandenbrouck Y, Viari A, et al. GenoLink: A graph-based querying and browsing system for investigating the function of genes and proteins. BMC Bioinformatics. 2006;7.
42. Vilela J, Asif M, Marques AR, Santos JX, Rasga C, Vicente A, et al. Biomedical knowledge graph embeddings for personalized medicine: Predicting disease-gene associations. Expert Syst. 2023;40(5).
43. Scherp A, Groener G, Škoda P, et al. Semantic web: past, present, and future. Trans Graph Data Knowl. 2024;2(1). doi:10.48550/arXiv.2412.17159.
44. Celebi R, Uyar H, Yasar E, Gumus O, Dikenelli O, Dumontier M. Evaluation of knowledge graph embedding approaches for drug-drug interaction prediction in realistic settings. BMC Bioinformatics. 2019;20(1).
45. Wang M, Qiu L, Wang X. A survey on knowledge graph embeddings for link prediction. Symmetry (Basel). 2021;13(3).
46. Murakami K, Tago SI, Takishita S, Morikawa H, Kojima R, Yokoyama K, et al. Pathogenicity Prediction of Gene Fusion in Structural Variations: A Knowledge Graph-Infused Explainable Artificial Intelligence (XAI) Framework. Cancers (Basel). 2024;16(10).
47. Chandak P, Huang K, Zitnik M. Building a knowledge graph to enable precision medicine. Sci Data. 2023;10(1).
48. Rajabi E, Etminani K. Knowledge-graph-based explainable AI: A systematic review. J Inf Sci. 2024;50(4):1019–29.

49. Azzi R, Bordea G, Griffier R, et al. Enriching the FIDEO ontology with food-drug interactions from online knowledge sources. J Biomed Semantics. 2024;15:3. doi:10.1186/s13326-024-00302-5.
50. Feng F, Tang F, Gao Y, Zhu D, Li T, Yang S, et al. GenomicKB: a knowledge graph for the human genome. Nucleic Acids Res. 2022;51(1 D):D950–6.
51. Chatzianastasis M, Vazirgiannis M, Zhang Z. Explainable Multilayer Graph Neural Network for cancer gene prediction. Bioinformatics. 2023;39(11).
52. Abe S, Tago S, Yokoyama K, Ogawa M, Takei T, Imoto S, et al. Explainable AI for Estimating Pathogenicity of Genetic Variants Using Large-Scale Knowledge Graphs. Cancers (Basel). 2023;15(4).
53. Zafar MR, Khan N. Deterministic Local Interpretable Model-Agnostic Explanations for Stable Explainability. Mach Learn Knowl Extr. 2021;3(3):525–41.
54. Aisopos F, Paliouras G. Comparing methods for drug–gene interaction prediction on the biomedical literature knowledge graph: performance versus explainability. BMC Bioinformatics. 2023;24(1).
55. Ahmed M, Dar AR, Helfert M, Khan A, Kim J. Data Provenance in Healthcare: Approaches, Challenges, and Future Directions. Sensors. 2023;23(14).
56. Roscher R, Bohn B, Duarte MF, Garcke J. Explainable Machine Learning for Scientific Insights and Discoveries. IEEE Access. 2020;8:42200–16.
57. Dibaeinia P, Ojha A, Sinha S. Interpretable AI for inference of causal molecular relationships from omics data. Sci Adv . 2025;11(7).
58. Rigden DJ, Fernández XM. The 2024 Nucleic Acids Research database issue and the online molecular biology database collection. Nucleic Acids Res. 2024;52(D1):D1-D9. doi:10.1093/nar/gkad1173.
59. Karami H, Derakhshani A, Ghasemigol M, et al. Weighted gene co-expression network analysis combined with machine learning validation to identify key modules and hub genes associated with SARS-CoV-2 infection. J Clin Med. 2021;10:3567.
60. Yang Y, Lu Y, Yan W. A comprehensive review on knowledge graphs for complex diseases. Brief Bioinform. 2023;24(1).
61. Althagafi A, Zhapa-Camacho F, Hoehndorf R. Prioritizing genomic variants through neuro-symbolic, knowledge-enhanced learning. Bioinformatics. 2024;40(5).
62. Irshad O, Ghani U. A Comparative Analysis of Biological Data Integration Systems Famous for Data Exploitation and Knowledge Discovery. Current Bioinformatics 2021; 16(5): pp. 662 - 681. DOI: 10.2174/1574893615999210101125442.
63. Irshad O, Ghani U, Iqbal R, et al. Performance optimization of IoT based biological systems using deep learning. Computer Communications 2020; 155: pp. 24-31. https://doi.org/10.1016/j.comcom.2020.02.059.