Searching the Druggable Genome using Large Language Models

Lars Schimmelpfennig[1], Matthew Cannon[2], Quentin Cody[3], Joshua McMichael[4], Adam C. Coffman[1], Susanna Kiwala[1], Kilannin Krysiak[5], Alex H. Wagner[2,6†], Malachi Griffith[1,4,7†], Obi L. Griffith[1,4,7†]

[1] Department of Medicine, Washington University School of Medicine, St. Louis, Missouri, USA

[2] Institute for Genomic Medicine (IGM), Nationwide Children's Hospital, Columbus, OH

[3] Independent Researcher, Philadelphia, PA, USA

[4] McDonnell Genome Institute, Washington University School of Medicine, St. Louis, Missouri, USA

[5] Department of Pathology, Washington University School of Medicine, St. Louis, Missouri, USA

[6] Department of Pediatrics, The Ohio State University College of Medicine, Columbus, OH

[7] Siteman Cancer Center, Washington University School of Medicine, St. Louis, Missouri, USA

Correspondence:

[†]Obi L Griffith, PhD, Washington University School of Medicine, Room MCC10116, 4590 Childrens Pl, Saint Louis, MO 63110. Phone: (314) 747-9248. Email: obigriffith@wustl.edu

[†]Malachi Griffith, PhD, Washington University School of Medicine, Room MCC10115, 4590 Childrens Pl, Saint Louis, MO 63110. Phone: (314) 286-1274. Email: mgriffit@wustl.edu

[†]Alex Wagner, PhD, Nationwide Children's Hospital, Room WB3155, Research Building 3, 575 Children's Crossroad, Columbus, OH 43210. Phone: (614) 355-1645, Email: Alex.Wagner@nationwidechildrens.org

ABSTRACT

Summary: The druggable genome encompasses the genes that are known or predicted to interact with drugs. The Drug-Gene Interaction Database (DGIdb) provides an integrated resource for discovering and contextualizing these interactions, supporting a broad range of research and clinical applications. DGIdb is currently accessed through structured web interfaces and API calls, requiring users to translate natural-language questions into database-specific query patterns. To allow for the use of DGIdb through natural language, we developed the DGIdb Model Context Protocol (MCP) server, which allows large language models (LLMs) access to up-to-date information through the DGIdb API. We demonstrate that the MCP server greatly enhances an LLM's ability to answer questions requiring accurate, up-to-date biomedical knowledge drawn from structured external resources.

Availability and implementation: The DGIdb MCP server is detailed at https://github.com/griffithlab/dgidb-mcp-server and includes instructions for accessing the server through the Claude desktop app.

1. Introduction

Drug-gene interactions play a central role in precision medicine, as they help determine which treatments are likely to be effective for a given patient and which may lead to resistance or adverse effects. As large-scale molecular studies increasingly reveal large sets of genes associated with disease, researchers need tools that can connect these findings to potential therapeutic strategies. The Drug-Gene Interaction Database (DGIdb)[1–5] addresses this need by aggregating and harmonizing drug-gene interaction information from multiple sources, providing a centralized resource for assessing gene druggability and therapeutic relevance. DGIdb supports downstream research and clinical-informatics applications, including gene-drug linking for prioritized disease genes[6], exploratory hypothesis generation[7], drug repurposing[8], genomic gene-list annotation[9], and precision oncology workflows[10]. The value of searching a catalog of drug-gene interactions is well illustrated by the challenge of *FLT3-mutated Acute Myeloid Leukemia*. Although the FLT3 inhibitor *Gilteritinib* is initially effective, patients often develop resistance due to additional genetic changes that emerge during treatment.[11] In such settings, DGIdb could be used to identify drugs known or predicted to interact with genes implicated in resistance, supporting exploratory analyses and prioritization of alternative therapeutic candidates.

Despite DGIdb's utility, this process can be time-consuming in multistep analyses, such as studies of acquired drug resistance, where users must first identify resistance-associated genes, then perform separate queries for each gene, and finally manually combine and prioritize potentially large lists of interacting drugs. As a result, DGIdb is difficult to use directly within natural-language workflows and is largely inaccessible to large language models (LLMs), which cannot natively construct or execute these structured queries. At the same time, LLMs have emerged as promising tools for a wide range of biomedical applications, including literature summarization and clinical decision support, due to their ability to synthesize complex information and respond flexibly to natural-language questions.[12–16] However, LLMs rely on static internal knowledge, may generate inaccurate or unsupported statements, and cannot directly access

65 curated biomedical resources without additional infrastructure.[17] These limitations restrict their
66 reliability in settings where accuracy and traceability are essential.

67 To address these challenges, the Model Context Protocol (MCP) provides a standardized
68 way to connect LLMs to authoritative external data sources. MCP servers serve as a bridge
69 between the flexibility of LLMs and the precision of curated biomedical databases and have been
70 increasingly applied to biomedical domain APIs, databases, and literature search endpoints,
71 including variant-level knowledgebases via the CIViC MCP server.[18–20] The DGIdb MCP server
72 builds upon this framework, allowing LLMs and researchers to retrieve, interpret, and cite drug-
73 gene interaction information directly from the curated resources, making it easier to integrate
74 accurate interaction data into genome-wide analyses and therapeutic decision-making.

75 **2. DGIdb MCP Server Description**

76 The DGIdb MCP server is publicly available through Cloudflare. The server provides four
77 tools, each corresponding to predefined GraphQL queries for integration with LLMs. These tools
78 allow LLMs to retrieve drug information, gene information, and drug-gene interactions based on
79 either a supplied gene list or drug list. The drug information tool returns key attributes such as
80 FDA approval status, drug classifications including immunotherapy and antineoplastic drugs, and
81 additional characteristics such as drug class. The gene information tool provides DGIdb gene
82 category annotations, such as kinase, ligase, and nuclear receptor designations, along with other
83 attributes relevant to gene druggability. Together, these data sources supply structured and well-
84 curated information that supplements an LLM's understanding of the drug-gene interaction space.
85 When the LLM calls the corresponding interaction query with a supplied list of genes or drugs,
86 the server returns a ranked list of their known interactions. Each interaction includes an interaction
87 type and, when applicable, directionality (activating or inhibitory) which together describe the
88 mechanism of action and effect on biological activity. Every interaction is also linked to its
89 supporting evidence, which may come from a publication or a curated database. Interactions are
90 ranked primarily according to the FDA approval status of the associated drugs, and secondarily

91 according to the DGIdb interaction score, which was introduced in the DGIdb 4.0 release.[4] The
92 score combines evidence strength (number of supporting publications and curated sources) with
93 interaction specificity, calculated from the relative number of known partners for the drug and
94 gene compared to global averages. The final score prioritizes interactions that are both well
95 supported and more specific, and remains static within a data release. This ranking strategy
96 allows LLMs to prioritize clinically informative and well-supported interactions.

97 To accommodate spelling variations and alternate names, the MCP server normalizes
98 drug and gene inputs to the closest matching label recognized by DGIdb. Alias similarity is scored
99 using Dice's coefficient over bigrams, which requires limited memory making it ideal for Cloudflare
100 workers. The MCP server obtains drug and gene aliases from the VICC Normalization Services,
101 which provide a standardized set of identifiers and synonyms for use during the scoring
102 process.[21,22]

103 **3. Example Use Cases**

104 1. A user might ask, "Which FDA-approved drugs target the gene *KIT*, and what is the
105 supporting evidence?" In response, the LLM identifies the gene of interest and the user's intent,
106 then invokes the appropriate MCP tool. The MCP server normalizes the gene name and issues
107 the corresponding GraphQL query to the DGIdb API. It returns a ranked list of drugs that interact
108 with *KIT*, along with interaction metadata such as interaction score, interaction type and direction,
109 FDA-approval status, and supporting sources. Using this structured output, the LLM composes a
110 user-facing answer that includes relevant evidence and links to the cited publications. A more
111 complete list of use cases for the DGIdb MCP server are outlined in Supplementary Table 1.

112 2. A second example use case demonstrates the joint use of the CIViC MCP[20] and DGIdb
113 MCP servers for drug candidate selection (Figure 1). A user may ask, "What genes can cause
114 resistance to *Ibrutinib* in *Chronic Lymphocytic Leukemia*, and what alternative drugs can target
115 them?" The LLM identifies the relevant CIViC entities (Therapy, Disease, Clinical Significance)
116 and invokes the appropriate tool from the CIViC MCP server to retrieve variant evidence. The

117 CIViC MCP server issues a GraphQL query to the CIViC API, which returns structured Evidence
118 Items associated with *Ibrutinib* resistance in *CLL*. From these records, the LLM extracts *BTK* as
119 the primary gene implicated in resistance and composes a query to the DGIdb MCP server to
120 identify other drugs that interact with *BTK*. The server returns a ranked list of drug-gene
121 interactions, enabling the LLM to respond with the highest-priority drug candidates, including
122 *Tirabrutinib, Acalabrutinib,* and *Zanubrutinib*. Using the outputs from each MCP server, the LLM
123 returns an answer that cites relevant curated evidence and publications. Additional examples of
124 how the DGIdb and CIViC MCP servers can be used together are outlined in Supplementary
125 Table 2.

126 **4. Benchmarking LLM Performance with DGIdb and CIViC MCP Servers**

127 We developed a set of classification and information-retrieval tasks to measure how
128 effectively the MCP server enhances an LLM's ability to answer questions requiring accurate and
129 current biomedical knowledge. As a first task, we asked GPT-5 to label drugs according to three
130 DGIdb-derived attributes: FDA approval status, immunotherapy classification, and antineoplastic
131 classification (Supplementary Table 3). We selected 100 drug names using stratified sampling to
132 approximate a 50/50 balance of positive and negative examples for each label. Because relatively
133 few drugs in DGIdb are classified as immunotherapies, achieving a perfectly even split was not
134 possible. GPT-5 was prompted to return three comma-separated YES/NO values to facilitate
135 parsing (Supplementary Prompt 1). Across all labels, GPT-5 augmented with the DGIdb MCP
136 substantially outperformed GPT-5 alone achieving a macro weighted F1-score of 0.99 compared
137 to 0.75. GPT-5 without MCP struggled particularly with identifying immunotherapies, producing
138 many false negatives and yielding a recall of only 0.38.

139 Next, we compared how often GPT-5 would invoke the DGIdb MCP server when
140 requesting the same drug information without explicitly mentioning DGIdb (Supplementary
141 Prompt 2). The DGIdb MCP server was available to the model in the same configuration as the
142 previous task, but tool use was not explicitly instructed. GPT-5 did not use the DGIdb MCP server

143 for 55 out of the 100 drugs. Among cases where GPT-5 did not invoke the MCP server,
144 performance declined, particularly for immunotherapy and antineoplastic classification. Recall
145 dropped from 1.00 to 0.68 for immunotherapies and from 0.88 to 0.68 for antineoplastic drugs
146 (Supplementary Table 4). To understand when GPT-5 decided to use the MCP server, we
147 manually inspected a random subset of predictions. In these examples, the MCP server was
148 generally not invoked for widely used, FDA-approved therapeutics with unambiguous, commonly
149 referenced names (e.g. Tamoxifen), but was preferentially used for uncommon, structurally
150 complex molecules, or database-style identifiers (e.g. ODM-203), indicating that GPT-5
151 selectively employed the MCP server when additional disambiguation or external grounding was
152 likely required.

153 Lastly, we evaluated a task that used both the CIViC MCP and DGIdb MCP servers to
154 demonstrate the potential of LLM-assisted drug candidate selection, as outlined in Supplementary
155 Figure 1. To allow for direct evaluation of the intermediate gene list, this task utilized two prompts.
156 In the first step, GPT-5 was prompted to identify genes associated with resistance to a specified
157 therapy in a given disease using CIViC-derived information (Supplementary Prompt 3). In the
158 second step, we issued a separate prompt for each gene identified in the first step, asking GPT-
159 5 to list drugs that interact with that gene using DGIdb-derived information (Supplementary
160 Prompt 4). To keep runtime manageable, we limited the request to the top 20 interacting drugs
161 per gene, as full interaction lists for a single gene can exceed 500 entries. For this evaluation, we
162 randomly selected 50 disease-therapy combinations from CIViC evidence as starting points for
163 the drug-candidate identification process. GPT-5 was instructed to rank genes according to the
164 strength of their associated CIViC evidence. Genes were ordered first by Evidence Level, with
165 level A considered the strongest and level E the weakest, and then by Evidence Rating, prioritizing
166 higher ratings (5 to 1). Drugs were ranked primarily by FDA approval status and secondarily by
167 interaction score. Because the multihop task requires GPT-5 to return genes and drugs in a
168 prioritized order, we evaluate correctness using precision, recall, and F1, and we assess ranking

169 quality using normalized discounted cumulative gain (NDCG). NDCG is a common information
170 retrieval metric that measures how well the model ranks relevant items, giving greater weight to
171 items that appear earlier in the list. NDCG ranges from 0 to 1, with higher values indicating
172 rankings that more closely match the ideal ordering. GPT-5 augmented with the CIViC and DGIdb
173 MCP servers substantially outperformed GPT-5 alone on the drug-list task, improving F1 from
174 0.14 to 0.95 and NDCG from 0.19 to 0.93 (Supplementary Table 5).

175 **5. Conclusion**

176 We demonstrate that the DGIdb MCP server enables reliable, structured exploration of
177 drug-gene interactions by allowing an LLM to query a comprehensive, curated interaction
178 resource in response to natural-language questions. By grounding GPT-5 in DGIdb, this approach
179 supports accurate retrieval, ranking, and contextualization of interaction evidence, which is
180 particularly important for tasks such as identifying alternative therapeutic candidates and
181 interpreting resistance-associated genes.

182 We further show that LLMs can make effective use of multiple MCP servers
183 simultaneously. In a drug-candidate identification task for therapy resistance, GPT-5 successfully
184 combined variant-level evidence retrieved from the CIViC MCP server with drug-gene interaction
185 data from the DGIdb MCP server, demonstrating how MCP servers can be composed to support
186 multi-step biomedical reasoning workflows that require information from multiple sources.

187 Our evaluation also highlights an important practical consideration: GPT-5 did not
188 consistently invoke MCP servers unless the relevant biomedical resource is mentioned explicitly
189 in the prompt. When the source (e.g., “DGIdb”) is named, GPT-5 reliably delegates retrieval to
190 the MCP server. When it is omitted, however, the model often defaults to its internal knowledge,
191 resulting in reduced accuracy, particularly for identifying immunotherapies. This finding
192 emphasizes the importance of prompt design in ensuring consistent MCP usage.

193 While the MCP framework has previously been applied to variant-level knowledge through
194 the CIViC MCP server[20], this work extends the paradigm to the drug-gene interaction domain and

195 illustrates how multiple MCP servers can be integrated to address more complex biomedical
196 questions. As future work, the development of standardized benchmarks for MCP-assisted
197 biomedical tasks, such as drug-gene interaction retrieval, variant interpretation, and cross-
198 knowledgebase reasoning, will be critical for enabling consistent evaluation and comparison of
199 MCP-enabled LLM systems.

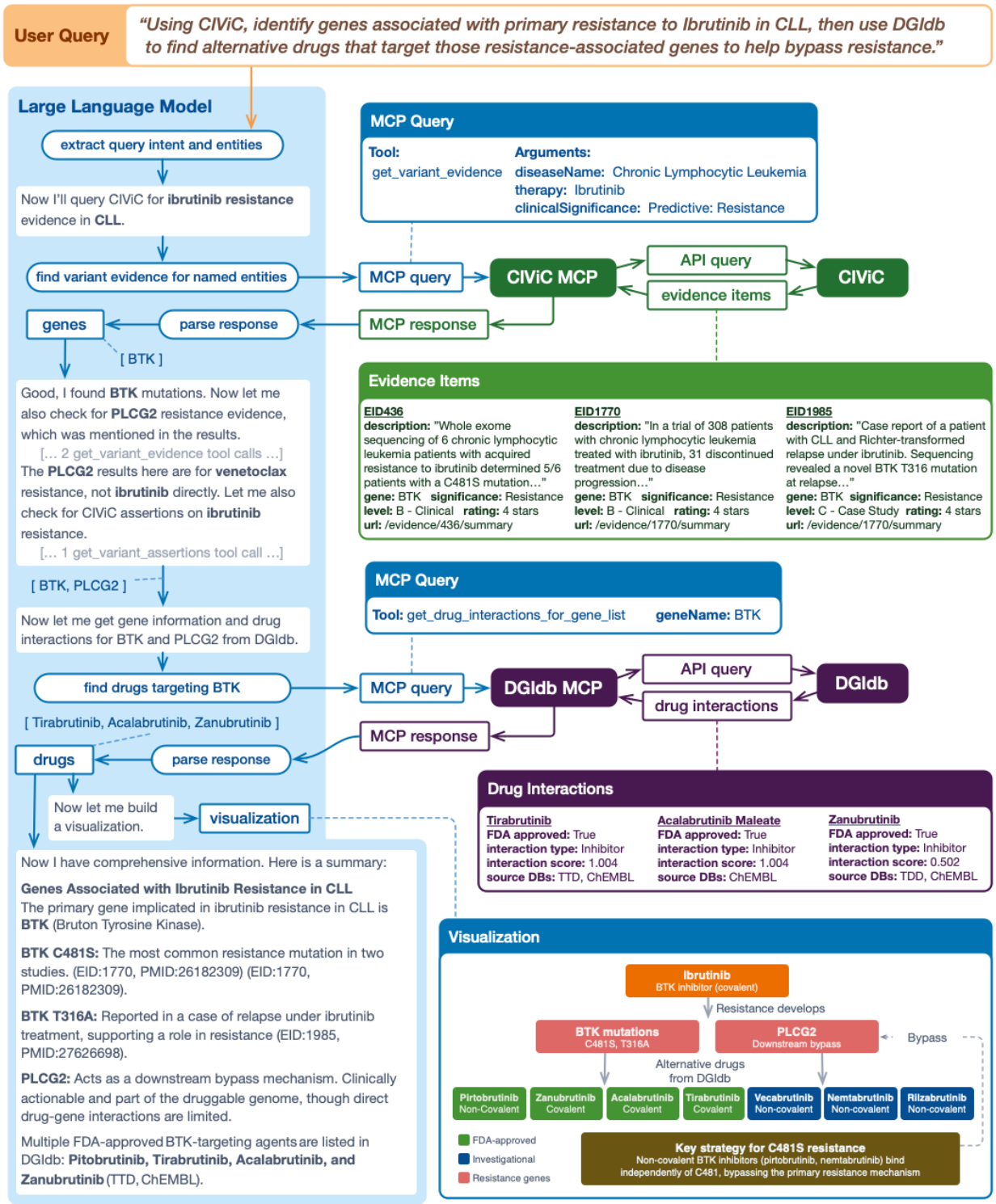

Figure 1. Joint use of the CIViC MCP and DGIdb MCP servers for drug candidate selection with Claude. A user asks, "What genes can cause resistance to Ibrutinib in Chronic Lymphocytic Leukemia, and what alternative drugs can target them?" The LLM: (1) extracts the users intent

204 and relevant entities (Therapy, Disease, Significance), (2) composes a query to the CIViC MCP
205 server with the appropriate tool selection, (3) the CIViC MCP server queries the CIViC API and
206 returns Evidence Items, (4) the LLM parses the MCP response and identifies the BTK as an
207 ibrutinib resistance gene, (5) the LLM composes a new query to the DGIdb MCP server to identify
208 other drugs that interact with BTK, (6) the DGIdb MCP server queries the DGIdb API and returns
209 drug interactions for BTK, (7) the LLM composes a summary.
210

211 **CONFLICT OF INTEREST**

212 None declared.
213

214 **FUNDING**

215 This work was supported by the National Institutes of Health (NIH), including the National Cancer
216 Institute (NCI) under award numbers U24CA305456, U24CA237719, U24CA258115, and
217 U24CA275783, and the National Human Genome Research Institute (NHGRI) under award
218 number R35HG011949.
219

220 **REFERENCES**

221 1. Griffith M, Griffith OL, Coffman AC, et al. DGIdb: mining the druggable genome. *Nat*
222 *Methods*. 2013;10(12):1209-1210. doi:10.1038/nmeth.2689
223 2. Wagner AH, Coffman AC, Ainscough BJ, et al. DGIdb 2.0: mining clinically relevant drug-
224 gene interactions. *Nucleic Acids Res*. 2016;44(D1):D1036-1044. doi:10.1093/nar/gkv1165
225 3. Cotto KC, Wagner AH, Feng YY, et al. DGIdb 3.0: a redesign and expansion of the drug-
226 gene interaction database. *Nucleic Acids Res*. 2018;46(D1):D1068-D1073.
227 doi:10.1093/nar/gkx1143

228 4. Freshour SL, Kiwala S, Cotto KC, et al. Integration of the Drug-Gene Interaction Database
229 (DGIdb 4.0) with open crowdsource efforts. *Nucleic Acids Res*. 2021;49(D1):D1144-D1151.
230 doi:10.1093/nar/gkaa1084
231 5. Cannon M, Stevenson J, Stahl K, et al. DGIdb 5.0: rebuilding the drug-gene interaction
232 database for precision medicine and drug discovery platforms. *Nucleic Acids Res*.
233 2024;52(D1):D1227-D1235. doi:10.1093/nar/gkad1040
234 6. Pranata R, Ratnasari N, Bayupurnama P, et al. Genomic analysis unlocks the potential of
235 denosumab as a targeted therapy for Crohn's disease. *Sci Rep*. 2025;15(1):44713.
236 doi:10.1038/s41598-025-28395-7
237 7. Gupta S, Srivastava P, Daniel R, et al. Germline activating sequence variations in
238 RASopathy spectrum genes: genotype–phenotype correlation in a North Indian cohort. *Front*
239 *Genet*. 2025;16. doi:10.3389/fgene.2025.1677143
240 8. Torricelli F, Sauta E, Manicardi V, et al. An Innovative Drug Repurposing Approach to
241 Restrain Endometrial Cancer Metastatization. *Cells*. 2023;12(5):794.
242 doi:10.3390/cells12050794
243 9. Lamisa AB, Ahammad I, Bhattacharjee A, et al. A meta-analysis of bulk RNA-seq datasets
244 identifies potential biomarkers and repurposable therapeutics against Alzheimer's disease. *Sci*
245 *Rep*. 2024;14(1):24717. doi:10.1038/s41598-024-75431-z
246 10. Cha S, Lee J, Shin JY, et al. Clinical application of genomic profiling to find druggable
247 targets for adolescent and young adult (AYA) cancer patients with metastasis. *BMC Cancer*.
248 2016;16:170. doi:10.1186/s12885-016-2209-1
249 11. McMahon CM, Ferng T, Canaani J, et al. Clonal Selection with RAS Pathway Activation
250 Mediates Secondary Clinical Resistance to Selective FLT3 Inhibition in Acute Myeloid
251 Leukemia. *Cancer Discov*. 2019;9(8):1050-1063. doi:10.1158/2159-8290.CD-18-1453
252 12. Wang Z, Cao L, Danek B, Jin Q, Lu Z, Sun J. Accelerating clinical evidence synthesis with
253 large language models. *Npj Digit Med*. 2025;8(1):509. doi:10.1038/s41746-025-01840-7

254 13. Chen Q, Hu Y, Peng X, et al. Benchmarking large language models for biomedical natural
255 language processing applications and recommendations. *Nat Commun*. 2025;16:3280.
256 doi:10.1038/s41467-025-56989-2
257 14. Williams CYK, Miao BY, Kornblith AE, Butte AJ. Evaluating the use of large language
258 models to provide clinical recommendations in the Emergency Department. *Nat Commun*.
259 2024;15(1):8236. doi:10.1038/s41467-024-52415-1
260 15. Hager P, Jungmann F, Holland R, et al. Evaluation and mitigation of the limitations of large
261 language models in clinical decision-making. *Nat Med*. 2024;30(9):2613-2622.
262 doi:10.1038/s41591-024-03097-1
263 16. Li Y, Zhao J, Li M, et al. RefAI: a GPT-powered retrieval-augmented generative tool for
264 biomedical literature recommendation and summarization. *J Am Med Inform Assoc JAMIA*.
265 2024;31(9):2030-2039. doi:10.1093/jamia/ocae129
266 17. Bhattacharyya M, Miller VM, Bhattacharyya D, Miller LE. High Rates of Fabricated and
267 Inaccurate References in ChatGPT-Generated Medical Content. *Cureus*. 2023;15(5):e39238.
268 doi:10.7759/cureus.39238
269 18. Ding P, Brettin T, Stevens R. BioR5: A Three-Layer Architecture for Biological Reasoning
270 in Scientific AI. In: *Proceedings of the SC '25 Workshops of the International Conference for*
271 *High Performance Computing, Networking, Storage and Analysis*. SC Workshops '25.
272 Association for Computing Machinery; 2025:594-601. doi:10.1145/3731599.3767400
273 19. Türei D, Schaul J, Palacio-Escat N, et al. OmniPath: integrated knowledgebase for multi-
274 omics analysis. *Nucleic Acids Res*. Published online November 18, 2025:gkaf1126.
275 doi:10.1093/nar/gkaf1126
276 20. Schimmelpfennig L, Cody Q, McMichael J, et al. CIViC MCP: Integrating Large Language
277 Models with the Clinical Interpretations of Variants in Cancer. *bioRxiv*. Preprint posted online
278 October 16, 2025:2025.10.13.682185. doi:10.1101/2025.10.13.682185

279 21. Cannon M, Stevenson J, Kuzma K, et al. Normalization of drug and therapeutic concepts
280 with Thera-Py. *JAMIA Open*. 2023;6(4):ooad093. doi:10.1093/jamiaopen/ooad093
281 22. Kuzma K, Stevenson J, Wagner A. VICC Gene Normalization Service. Published online
282 July 21, 2025. doi:10.5281/zenodo.16272753