

# 1 TICTAC: Target Illumination Clinical Trial Analytics with Cheminformatics

2 Jeremiah I Abok<sup>1</sup>, Jeremy S Edwards<sup>1</sup>, Jeremy J Yang<sup>2\*</sup>

3 <sup>1</sup>Department of Chemistry & Chemical Biology, University of New Mexico, Albuquerque, NM,  
4 87106 USA

5 <sup>2</sup>Division of Translational Informatics, Department of Internal Medicine, University of New Mexico  
6 Health Sciences Center, Albuquerque, NM 87131, USA

7 \* **Correspondence:**

8 Corresponding Author

9 [jjyang@salud.unm.edu](mailto:jjyang@salud.unm.edu)

10 **Keywords:** Disease-Target<sub>1</sub>, Drug discovery<sub>2</sub>, Hypothesis generation<sub>3</sub>, Clinical trial data<sub>4</sub>,  
11 Inferences<sub>5</sub>.

## 12 Abstract

### 13 Introduction

14 Identifying disease–target associations is a pivotal step in drug discovery, offering insights that guide  
15 the development and optimization of therapeutic interventions. Clinical trial data serves as a valuable  
16 source for inferring these associations. However, issues such as inconsistent data quality and limited  
17 interpretability pose significant challenges. To overcome these limitations, an integrated approach is  
18 required that consolidates evidence from diverse data sources to support the effective prioritization of  
19 biological targets for further research.

### 20 Methods

21 We developed a comprehensive data integration and visualization pipeline to infer and evaluate  
22 associations between diseases and known and potential drug targets. This pipeline integrates clinical  
23 trial data with standardized metadata, providing an analytical workflow that enables the exploration of  
24 diseases linked to specific drug targets as well as facilitating the discovery of drug targets associated  
25 with specific diseases. The pipeline employs robust aggregation techniques to consolidate multivariate  
26 evidence from multiple studies, leveraging harmonized datasets to ensure consistency and reliability.  
27 Disease–target associations are systematically ranked and filtered using a rational scoring framework  
28 that assigns confidence scores derived from aggregated statistical metrics.

### 29 Results

30 Our pipeline evaluates disease–target associations by linking protein-coding genes to diseases and  
31 incorporates a confidence assessment method based on aggregated evidence. Metrics such as  
32 meanRank scores are employed to prioritize associations, enabling researchers to focus on the most  
33 promising hypotheses. This systematic approach streamlines the identification and prioritization of  
34 biological targets, enhancing hypothesis generation and evidence-based decision-making.

### 35 Discussion

36 This innovative pipeline provides a scalable solution for hypothesis generation, scoring, and ranking  
37 in drug discovery. As an open-source tool, it is equipped with publicly available datasets and designed  
38 for ease of use by researchers. The platform empowers scientists to make data-driven decisions in the  
39 prioritization of biological targets, facilitating the discovery of novel therapeutic opportunities.

40

## 41 **1 Introduction**

42 ClinicalTrials.gov, managed by the U.S. National Library of Medicine (NLM), is a critical public  
43 repository that enhances transparency and accessibility in biomedical research. Established in 2000  
44 under the Food and Drug Administration Modernization Act (FDAMA) of 1997, it serves as a  
45 centralized platform for registering clinical studies and disseminating trial results, addressing  
46 challenges related to data accessibility and reliability (Zarin et al., 2005; Tse et al., 2018). By  
47 supporting clinical trial registration, the database provides researchers, policymakers, and the public  
48 with comprehensive information about study designs, methodologies, and outcomes. Adherence to  
49 global reporting requirements, including those mandated by the International Committee of Medical  
50 Journal Editors (ICMJE) and the World Health Organization (WHO), ensures standardized practices  
51 across the research community (Laine et al., 2007). ClinicalTrials.gov aligns with FAIR (Findable,  
52 Accessible, Interoperable, and Reusable) principles (Wilkinson et al., 2016), employing standardized  
53 data formats to enhance interoperability across platforms and disciplines. It curates metadata,  
54 harmonizes submissions, and implements consistent reporting standards, addressing challenges such  
55 as incomplete reporting, variability in trial methodologies, and inconsistencies in diagnostic criteria  
56 (Riveros et al., 2013). The platform supports centralized trial registration and compliance with legal  
57 and ethical mandates for transparency (Zarin et al., 2011). Its study results database enables researchers  
58 to submit and access summary results, fostering evidence-based decision-making and mitigating  
59 publication bias through the inclusion of unpublished trial data (Prayle et al., 2012). Furthermore,  
60 integration with global registries, such as the WHO's International Clinical Trials Registry Platform  
61 (ICTRP), promotes harmonization and accessibility of clinical research data worldwide. Despite its  
62 significant contributions, ClinicalTrials.gov faces challenges including variations in reporting quality,  
63 delays in result submissions, and inconsistencies in terminology (DeVito et al., 2020). The platform  
64 continues to evolve by introducing advanced data validation tools, promoting adherence to global  
65 reporting standards, and collaborating with stakeholders to refine metadata frameworks and bridge  
66 reporting gaps. As a robust and scalable platform for managing clinical trial data, ClinicalTrials.gov  
67 fosters transparency and enhances research reproducibility while empowering the scientific community  
68 and improving public trust.

69 The Database for Aggregate Analysis of ClinicalTrials.gov (AACT) was introduced to address  
70 challenges in analyzing aggregate data from ClinicalTrials.gov, such as inconsistent data structures,  
71 variability in nomenclature, and evolving data collection practices. Its purpose is to enhance the  
72 usability of ClinicalTrials.gov data by consolidating and normalizing information, enabling more  
73 effective aggregate analysis, policy studies, and systematic evaluations of clinical trial attributes and  
74 trends (Tasneem et al., 2012). AACT transforms raw clinical trial data into a structured, enriched, and  
75 analyzable format, integrating Medical Subject Headings (MeSH) terms and advanced curation  
76 techniques to ensure consistency and usability (Tasneem et al., 2012). By enabling systematic reviews,  
77 policy analysis, and diverse applications, AACT plays a pivotal role in evaluating global trial trends,  
78 aligning with FAIR principles, and driving innovation in clinical research. Despite challenges like  
79 reliance on MeSH hierarchies and limited global representation, ongoing advancements in ontology

80 integration and interoperability aim to position AACT as an indispensable resource for the future of  
81 clinical trial analysis.

82

83 The core principle of pharmacology is that drugs, whether small molecules or biologics are designed  
84 to specifically interact with target molecules, often proteins, to modulate physiological processes and  
85 influence disease progression (Moffat et al., 2017; Scannell et al., 2012). Advanced methods in the  
86 pharmaceutical industry facilitate the discovery and optimization of these drugs, addressing challenges  
87 in efficacy, dosing, and safety for market approval (Hay et al., 2014). However, analysis of drug  
88 development pipelines reveals that insufficient efficacy, particularly in late-stage clinical trials, is a  
89 primary cause of failure, often due to inadequate validation of the target's role in disease physiology  
90 (Ledford, 2011; Liu et al., 2021). This highlights the need for rigorous evidence supporting target-  
91 disease associations to improve success rates and minimize costly late-stage failures (Dahlin et al.,  
92 2015). Traditionally, drug targets were selected based on experimental evidence linking their  
93 modulation to disease outcomes (Muller & Milton, 2012). Recent advancements in high-throughput  
94 technologies, such as sequencing, genotyping, and mass spectrometry, have enhanced our ability to  
95 characterize biological samples, uncovering new opportunities to understand disease mechanisms  
96 (Vincent et al., 2015; Huang et al., 2011). Furthermore, the growing repository of clinical trial data,  
97 alongside extensive literature, serves as a valuable resource for identifying targets and generating  
98 hypotheses to inform the drug discovery process (Lysenko et al., 2018).

99 Here, we present TICTAC (Target Illumination Clinical Trial Analytics with Cheminformatics), an  
100 application designed to illuminate understudied drug targets by leveraging aggregated data from  
101 AACT. TICTAC enables ranking, filtering, and interpretation of inferred disease–target associations,  
102 assigning scores derived from aggregated evidence linking diseases to protein-coding genes mapped  
103 from drugs. This study outlines the analytical framework and interpretability of TICTAC, addressing  
104 statistical and semantic challenges. TICTAC demonstrates the application of data science in achieving  
105 scientific consensus and improving interpretability. The validation of disease-target associations is  
106 critical for ensuring the accuracy and reliability of biomedical datasets. This study also presents a  
107 methodology for validating the TICTAC disease-gene associations against MedlineGenomics  
108 (formerly Genetics Home Reference) by leveraging standardized disease terminologies, such as  
109 Disease Ontology IDs (DOIDs) (Schriml et al., 2022) and UMLS Concept Unique Identifiers (CUIs)  
110 (McInnes et al., 2007). By comparing disease-target associations across these datasets, the aim is to  
111 quantify overlap, identify areas of divergence, and provide insights into the consistency and reliability  
112 of the two data sources.

113

## 114 **2 Material & Methods**

### 115 **2.1 AACT Data Preprocessing**

116 The initial step in our methodology involved applying NextMove LeadMine (version 3.14.1).  
117 LeadMine, developed by NextMove Software, is a commercial text mining tool that identifies and  
118 annotates chemical entities, protein targets, genes, diseases, species, and more using curated grammars  
119 and dictionaries with advanced capabilities like correcting misspelled terms with CaffeineFix  
120 technology and supports chemical entity recognition in multiple languages, including Chinese and

121 Japanese(Lowe & Sayle, 2015). This tool was utilized to analyze intervention names and study  
122 descriptions retrieved from ClinicalTrials.gov, enabling the identification of unique drug names and  
123 their corresponding SMILES (Simplified Molecular Input Line Entry System) representations.  
124 However, many terms encountered, such as "placebo," "test product," "medication," and  
125 "chemotherapy," lacked specific structural chemical information. For disease-named entity recognition  
126 (NER), we leveraged the JensenLab Tagger (Cook & Jensen, 2019). The JensenLab Tagger is a  
127 dictionary-based NER tool designed to identify and annotate entities such as genes, proteins, species,  
128 diseases, and other biomedical terms within text. It has been instrumental in text-mining applications,  
129 including extracting protein-protein interactions and annotating biomedical literature. The source code  
130 is available on GitHub: <https://github.com/larsjuhljensen/tagger>. This tool was applied to trial  
131 descriptions to identify and categorize diseases, resolving disease mentions to standardized terms in  
132 the Disease Ontology (DOID). This enhanced the consistency and granularity of disease data across  
133 sources. Subsequently, compound-target mapping was conducted to identify potential drug targets.  
134 Chemical entities were mapped to PubChem(Kim et al., 2021) using the PUG REST API with  
135 SMILES-based exact search and to ChEMBL(Zdrazil et al., 2024) using REST API queries via  
136 InChIKey. Biological targets were then mapped from ChEMBL bioassays and linked to the Integrated  
137 Disease-Target Knowledgebase (IDG-TCRD/Pharos)(Sheils et al., 2021) using UniProt IDs (The  
138 UniProt Consortium, 2023). This systematic mapping established relationships between chemical  
139 entities, their biological targets, and associated diseases, forming a foundation for data aggregation  
140 efforts.

141 Our dataset, extracted from the Aggregate Analysis of ClinicalTrials.gov (AACT) database as of  
142 September 30, 2024, comprised 507,584 studies, each identified by a unique NCT\_ID. These studies  
143 referenced 901,776 publications, associated with 632,153 PubMed IDs (PMIDs) and 127,455  
144 RESULT-type references, spanning 170,697 unique NCT\_IDs. For drug-related data, 133,760 unique  
145 drug names were linked to 365,878 unique intervention IDs. It is worth noting that individual NCT\_IDs  
146 often referenced multiple drugs, with synonymous naming conventions contributing to challenges in  
147 precise drug identification.

148 Our analysis focused exclusively on interventional drug studies, excluding observational studies. Of  
149 the total dataset, 388,958 (76.6%) studies were classified as interventional, with 177,780 designated as  
150 interventional drug studies, each linked to a unique NCT\_ID. To refine the chemical data, NextMove  
151 LeadMine was used to resolve drug names into standardized chemical structures via SMILES notation.  
152 This process identified 6,595 unique SMILES associated with 23,982 unique intervention IDs and  
153 19,662 unique drug names. The resulting dataset included 901,776 study references and 632,153  
154 PubMed IDs, offering a robust basis for linking identified entities in clinical trial data to biomedical  
155 literature.

156

## 157 **2.2 Computations for Disease-Target Associations**

### 158 **nStudyNewness**

159 The **nStudyNewness** metric quantifies the recency-weighted relevance of clinical studies associated  
160 with each disease-gene pair, reflecting the higher impact of newer studies. This is achieved using an  
161 exponential decay function that prioritizes recent studies.

162

$$w_i = 2^{e-ti/h}$$

163 where:

- 164 •  $t_i$  is the age of the study in years,
- 165 •  $h$  is the half-life determining the decay rate ( $h = 5$  years for
- 166 studies  $\leq 10$  years old and  $h=10$  years for older studies).

167 The total nStudyNewness for a disease-target pair is:

$$168 \quad \text{nStudyNewness} = \sum_{n=1}^N (w_i)$$

169 where  $N$  is the total number of studies associated with the disease-target pair.

170

### 171 **nPublicationWeighted**

172 The **nPublicationWeighted** metric assigns weights to publications based on their reference type,  
173 accounting for their varying levels of impact on disease-target associations. For each publication  $i$   
174 with reference type  $r_i$ , the weight  $w_i$  is defined as:

- 175 •  $w_i=1.0$ : If the publication reference type  $r_i$  is a **Result** (denoted by 0.0), it is given the highest
- 176 weight.
- 177 •  $w_i=0.5$ : If  $r_i$  is a **Background** reference (denoted by 1.0), it is assigned a medium weight.
- 178 •  $w_i=0.25$ : If  $r_i$  is a **Derived** reference (denoted by 2.0), it is given the lowest weight.

### 179 **Ranking Associations**

180 To facilitate the prioritization of disease-gene associations, rankings are computed based on multiple  
181 metrics, culminating in an overall score.

- 182 1. **Rank Computation:** For each metric (nDiseases, nDrug, nStud, nPub, nStudyNewness,  
183 nPublicationWeighted), the rank  $R_{m,j}$  for disease-target pair  $j$  is computed in descending  
184 order:

$$185 \quad R_{m,j} = \text{Rank}(m,j)$$

- 186 2. **Mean Rank:** The average rank across all metrics for a disease-target pair  $j$  is:

$$187 \quad \text{meanRank}_j = \frac{1}{K} \sum_{n=1}^K R_{m,j}$$

188 where  $K$  is the total number of metrics.

- 189 3. **Percentile Rank:** Percentile rank is calculated to normalize the meanRank values:

$$190 \quad \text{percentileMeanRank}_j = \frac{1}{N} \text{Rank}(\text{meanRank}_j \times 100)$$

191 where  $N$  is the total number of disease-target pairs.

- 192 4. **Mean Rank Score:** The final mean rank score, providing a scale from 0 to 100, is:

$$193 \quad \text{meanRankScore}_j = 100 - (\text{percentileMeanRank}_j)$$

194



195 This methodology ensures inferred disease-target associations are systematically ranked based on  
196 multiple evidence metrics, with the highest meanRankScore indicating the most promising targets for  
197 further research and development.

## 198 **2.3 Validation Method**

199 The validation methodology included key steps to map and compare disease-target associations  
200 between TICTAC and MedlinePlus Genetics (MedlineGenomics) (U.S. National Library of Medicine,  
201 n.d). The mapping from Disease Ontology was processed to link DOIDs with UMLS CUIs and validate  
202 data integrity through unique counts. MedlinePlus genetics conditions were formatted to align with  
203 these mappings, ensuring compatibility and accuracy. TICTAC disease-target associations were  
204 similarly prepared by standardizing identifiers, refining data fields, and verifying counts of key entities.  
205 Common CUIs between datasets were identified, and their corresponding associations were extracted  
206 for comparison. Overlaps between the datasets were evaluated by analyzing shared associations, with  
207 the percentage of MedlinePlus Genetics associations present in TICTAC calculated. The validation  
208 assessed dataset integrity, overlap in conditions (common CUIs), and overlap in disease-target  
209 associations to measure alignment and reliability.

210

## 211 **3 Results**

### 212 **3.1 The TICTAC Application**

213 TICTAC supports drug target identification by scoring and ranking associations between drug targets  
214 (protein-coding genes) and diseases. The TICTAC workflow aggregates and filters inferred clinical  
215 trial findings to generate actionable insights. These insights can be leveraged to enhance target  
216 prioritization through interactive visualizations and hit lists (Fig. 2), enabling users to identify the  
217 strongest, evidence-supported associations.

218 Hits in the TICTAC dashboard are ranked by **meanRankScore**, as described in Section 2. The  
219 scatterplot presents evidence (meanRankScore) on the X-axis versus publication count (nPub) on the  
220 Y-axis, visually representing disease-target associations. The workflow allows users to filter results  
221 using query parameters such as disease terms (e.g., "type 2 diabetes/insulin resistance," DOID:9352)  
222 and gene symbols. Data points in the scatterplot are colored by target development level (TDL), a  
223 knowledge-based classification system that categorizes human proteins into four distinct groups (Oprea  
224 et al., 2018) as Tclin (Santos et al., 2017), Tchem, Tbio, and Tdark for the comprehensiveness of  
225 exploration from clinical, chemical, and biological perspectives. This framework enables the  
226 prioritization of disease-target pairs based on their evidence levels and functional classifications,  
227 facilitating drug discovery efforts and identifying novel targets for further research.

228

### 229 **3.2 Agreement-Based Validation of TICTAC and MedlinePlus Genetics Datasets**

230 The agreement-based validation process revealed that the TICTAC dataset includes 2,243 unique  
231 Disease Ontology IDs (DOIDs) and 2,022 unique gene symbols, while the MedlinePlus Genetics  
232 dataset contains 1,216 conditions mapped to Concept Unique Identifiers (CUIs) and 2,142 unique  
233 CUIs. A total of 193 CUIs were shared between the two datasets, allowing for a comparative analysis  
234 of associations. For the shared CUIs, the TICTAC dataset encompassed 63,569 associations involving  
235 1,804 gene symbols, whereas the MedlinePlus Genetics dataset comprised 1,247 associations involving

236 967 gene symbols. Notably, 136 associations overlapped between the two datasets, accounting for  
237 10.91% of the MedlinePlus Genetics associations. This indicates that the TICTAC dataset contains a  
238 high proportion of novel associations, with an estimated 89% of associations being unique to TICTAC.  
239 These results highlight the potential of the TICTAC dataset to contribute significant novel insights,  
240 while also demonstrating consistency between the datasets for a subset of shared associations. This  
241 agreement-based approach underscores the complementary nature of these resources and the value of  
242 TICTAC for expanding the landscape of disease-target associations for further investigation.

243

### 244 3.3 Using TICTAC for Drug Target Illumination

#### 245 3.3.1 Exploring Understudied Genes in Type2 Diabetes/Insulin Resistance: Integrating Metrics 246 for Target Discovery Potential

247 Fig. 3 highlights examples of understudied (Tbio) genes associated with "type 2 diabetes/insulin  
248 resistance" within the TICTAC framework. The analysis focuses on genes classified as Tbio, indicating  
249 they are less characterized and have limited research attention compared to more 'established targets'.  
250 The table provides key metrics for each gene-disease pair, including the number of diseases  
251 (nDiseases), drugs (nDrug), studies (nStud), and publications (nPub), along with evidence  
252 prioritization scores such as nStudyNewness and meanRankScore. These metrics collectively illustrate  
253 the relative evidence strength and research activity for each gene-disease pair. The analysis reveals  
254 significant variability in the evidence supporting these genes. For instance, some genes, such as  
255 UGT1A10 (41 publications, meanRankScore = 97.87784), have substantial supporting evidence, while  
256 others, such as COQ8A (0 publications, meanRankScore = 46.33545), remain less explored in the  
257 literature. Importantly, all listed genes have some level of therapeutic association, as reflected in the  
258 nDrug column, which indicates the number of drugs linked to each gene. For example, MAP4K4 and  
259 RAC1 are linked to drugs such as Sorafenib and Dasatinib, while other genes, such as CISD2, are  
260 associated with broader therapeutic contexts. The nStudyNewness metric emphasizes the relevance of  
261 recent evidence for these gene-disease pairs, with genes like CISD2 demonstrating strong support from  
262 newer studies and achieving a high meanRankScore (98.92213). Such metrics underscore the potential  
263 for these Tbio genes to serve as new therapeutic targets for metabolic disorders. Genes with high  
264 meanRankScore and robust evidence, such as CISD2, may warrant further investigation in drug  
265 discovery efforts. Conversely, genes with minimal publications, such as COQ8A, represent  
266 opportunities for expanding research into their roles in type 2 diabetes and insulin resistance.

267

268 Overall, this analysis of Tbio genes in the context of type 2 diabetes/insulin resistance highlights the  
269 utility of TICTAC in identifying and prioritizing understudied targets for further research and  
270 therapeutic development. By integrating metrics such as meanRankScore, nPub, and nStudyNewness,  
271 the framework enables systematic exploration of gene-disease associations, providing actionable  
272 insights for drug discovery.

273

#### 274 3.3.2 Integrating Evidence for Disease-Target Associations: Insights from the TICTAC 275 Provenance Dashboard

276 Figure 4 presents the TICTAC Provenance Dashboard, which provides detailed reference data  
277 supporting the association between the MCR4 (Melanocortin receptor 4) gene and type 2  
278 diabetes/insulin resistance (DOID:9352). This dashboard aggregates references from clinical trials and  
279 publications, offering comprehensive insights into the evidence linking the gene to the disease. The  
280 framework for the provenance of this association includes several key elements. The **nct\_id** column  
281 lists unique identifiers for clinical trials, while the **reference\_type** column categorizes evidence as

282 BACKGROUND (contextual information), RESULT (direct findings), or DERIVED (secondary  
283 findings). The **pmid** column includes PubMed IDs linking to corresponding publications, and the  
284 **citation** column provides bibliographic details such as titles, authors, journals, and publication years.  
285 The results reveal a mix of evidence types. BACKGROUND references highlight physiological  
286 mechanisms like insulin sensitivity and glucose metabolism foundational to type 2 diabetes research.  
287 RESULT references contribute direct findings, such as variability in cyclooxygenase inhibition in  
288 aspirin studies, while DERIVED references explore broader evidence, such as anti-psychotropic  
289 medication usage. For example, Larsen et al. (2012) offer insights into diabetes mechanisms.

290

291 Sodium-glucose cotransporter 2 (SGLT2) inhibitors have emerged as a novel therapeutic approach for  
292 type 2 diabetes, particularly in improving cardiovascular outcomes. These inhibitors reduce myocardial  
293 glucose uptake by up to 57%, inducing metabolic shifts toward fatty acid utilization (Lauritsen et al.,  
294 2021). Additionally, they lower myocardial blood flow (MBF) by approximately 13%, potentially due  
295 to diuretic effects that reduce blood pressure and alter renal hemodynamics (Kimura, 2016). While  
296 SGLT2 inhibitors significantly mitigate glucose uptake and blood flow, their broader cardioprotective  
297 effects likely involve mechanisms beyond these metrics. For example, they do not notably alter  
298 myocardial oxygen consumption, suggesting additional pathways are at play (Lauritsen et al., 2021;  
299 Huang et al., 2023). Despite these complexities, improved cardiac function and reduced heart failure  
300 risk highlight their therapeutic potential (Huang et al., 2023). A strong hypothesis thus emerges:  
301 SGLT2 inhibition improves cardiac outcomes in individuals with type 2 diabetes by inducing metabolic  
302 shifts, reducing myocardial blood flow, and enhancing cardiac efficiency. This framework underscores  
303 the importance of integrating diverse evidence types to establish robust disease-target associations.  
304 Direct access to clinical trial identifiers and PubMed references enhances reproducibility and  
305 credibility, making the TICTAC Provenance Dashboard a valuable resource for researchers exploring  
306 therapeutic targets in type 2 diabetes.

### 307 **3.3.3 Mapping Disease-Target Associations and Evidence in Lung Cancer Research**

308 Figures 5 and 6 provide a comprehensive view of gene-disease associations related to lung cancer  
309 (DOID:1324), focusing on potential gene targets and the provenance of specific evidence linking the  
310 PGR (Progesterone receptor, UniProt: P06401) gene to the disease.

311

312 Figure 5 presents the TICTAC Disease-Target Dashboard, which highlights key metrics for genes  
313 associated with lung cancer. The dashboard includes metrics such as the number of diseases  
314 (nDiseases), drugs (nDrug), studies (nStud), and publications (nPub) related to each gene-disease pair.  
315 Additional evidence metrics such as nStudyNewness, which prioritizes recent studies, and  
316 PublicationWeight, which accounts for the relevance of evidence, are also shown. The meanRankScore  
317 provides a composite metric ranking associations based on the strength and quality of evidence. For  
318 each gene, the dashboard also lists the associated drugs and Target Development Level (TDL), a  
319 classification reflecting the druggability of the gene. Notable findings include highly ranked genes such  
320 as ADK (meanRankScore = 97.73276) and BDKRB2 (meanRankScore = 95.97179), which are  
321 supported by robust evidence from multiple studies, publications, and associated drugs.  
322 Therapeutically, drugs such as Abemaciclib, 5-Fluorouracil, and Carboplatin are frequently linked to  
323 the listed genes, emphasizing their relevance to lung cancer treatment. The genes in this dashboard are  
324 classified as Tchem, indicating their known potential as druggable targets, making this dashboard a



325 critical resource for prioritizing genes with strong evidence for further exploration in lung cancer  
326 research and therapeutic development.

327

328 Figure 6 provides detailed evidence provenance for the association between the PGR (Progesterone  
329 receptor) gene and lung cancer. The TICTAC Provenance Dashboard aggregates references from  
330 clinical trials and publications to trace the supporting evidence. It includes clinical trial identifiers  
331 (nct\_id), 2978 reference types categorized as RESULT for direct findings or BACKGROUND for  
332 supporting evidence, PubMed IDs (pmid), and detailed citations for each referenced study. The  
333 provenance dashboard reveals that the association between PGR and lung cancer is supported by both  
334 direct results and background studies. These references detail the role of therapies such as carboplatin  
335 in treating lung cancer, providing clinical context for PGR's relevance. This detailed provenance  
336 ensures the traceability of the data and supports further exploration of PGR in lung cancer research.

337

338 Taken together, Figures 5 and 6 illustrate the utility of TICTAC in mapping disease-target associations  
339 and integrating comprehensive evidence provenance. The TICTAC Disease-Target Dashboard enables  
340 the identification of high-priority gene targets and associated drugs, while the Provenance Dashboard  
341 provides detailed and traceable evidence for specific genes like PGR. This combination of tools allows  
342 for a robust prioritization framework, supporting translational research and advancing the discovery of  
343 potential therapeutic targets.

344

## 345 **4 Discussion**

### 346 **4.1 Illuminating Knowledge Gaps in Targets**

347 The National Institutes of Health's (NIH) Illuminating the Druggable Genome (IDG) program is  
348 dedicated to advancing our understanding of understudied proteins within pivotal druggable families,  
349 including G-protein-coupled receptors (GPCRs), ion channels, and protein kinases (National Institutes  
350 of Health, 2024). By elucidating the roles of these proteins in health and disease, the program seeks to  
351 identify novel therapeutic targets and foster innovative drug development strategies.

352

353 For a detailed exploration of the objectives and methodologies of the IDG program, refer to  
354 "Unexplored therapeutic opportunities in the human genome" by Oprea et al. (Oprea et al., 2018).  
355 TICTAC aligns seamlessly with this mission, focusing on evaluating clinical trial evidence to reveal  
356 disease-target associations. Unlike other platforms, such as Open Targets (Ghoussaini et al., 2021;  
357 Ochoa et al., 2021), which leverage a blend of data, supervised machine learning, and external sources,  
358 TICTAC strictly aggregates evidence from the AACT database. This approach minimizes biases  
359 inherent in curated training data and domain-specific assumptions, offering interpretable results that  
360 are grounded in experimental provenance and reproducible methodologies. By integrating its  
361 automated, sustainable workflow into resources such as the Pharos portal (Nguyen et al., 2017; Sheils  
362 et al., 2021), TICTAC complements existing tools for target illumination. While Open Targets supports  
363 scientists by enhancing associations with external validation, TICTAC prioritizes the rigor of direct

364 clinical trial data. This makes it particularly suited for downstream users who require traceable,  
365 transparent insights into disease–target relationships.

366

## 367 **4.2 Transforming Data into Actionable Knowledge**

368 In genomics and drug discovery, where data overload is a persistent challenge, transforming raw  
369 information into actionable knowledge remains critical. Specialized tools aid in uncovering insights,  
370 but the interpretation and integration of these findings often demand expertise that non-specialists may  
371 lack. TICTAC addresses this gap by introducing a layer of abstraction that is not only accessible but  
372 also directly aligned with the objectives of drug discovery scientists. TICTAC’s design philosophy  
373 rests on simplicity and scientific axioms, such as prioritizing evidence from independent confirmatory  
374 results (Yang et al., 2021). This emphasis ensures the interpretability of findings and highlights clinical  
375 trial results as a cornerstone of evidence-based reasoning. While correlation does not equate to  
376 causation, it builds a foundation of plausibility, fostering hypothesis generation and prioritization. With  
377 this context in mind, TICTAC provides a pragmatic, rational framework for ranking research  
378 hypotheses. Clinical trials are often influenced by experimental noise and systematic uncertainties  
379 stemming from factors like the COVID-19 pandemic’s impact on follow-ups (Servick, 2020),  
380 incomplete data from missing participants (Verzilli & Carpenter, 2002), and biases in endpoint  
381 evaluations (Chen et al., 2020). Ambiguities in defining study populations and inadequate reporting of  
382 sample sizes further complicate result interpretation (Frampton & Shepherd, 2008). Addressing these  
383 challenges through improved methodologies and reporting standards is essential to enhance the validity  
384 and reliability of trial outcomes. While acknowledging that experimental noise and systematic  
385 uncertainties can accompany clinical trial data, the approach ensures that aggregated insights are  
386 meaningful and usable for non-specialist stakeholders in drug development.

387

## 388 **4.3 Designing for Seamless Integration and Confidence**

389 Biomedical knowledge discovery thrives on the integration of diverse, heterogeneous data sources,  
390 reflecting the inherent complexity of the field. However, challenges related to provenance,  
391 interpretability, and confidence frequently undermine these efforts. TICTAC addresses these concerns  
392 by employing simple yet robust metrics, such as unbiased meanRank scores, to evaluate and rank  
393 disease–target associations. One of TICTAC’s key innovations lies in its transparent approach to  
394 confidence assessment. By restricting provenance to the linked publications, the platform ensures  
395 enhanced interpretability and traceability. This minimalist yet methodical approach reduces the  
396 accumulation of errors and confidence decay that often plague systems integrating multiple,  
397 heterogeneous data sources. Continuous confidence scores allow for dynamic algorithmic weighting  
398 and filtering, supporting downstream applications that require both precision and flexibility. The clear  
399 use of standardized identifiers and semantics strengthens integration across biomedical resources,  
400 paving the way for consistent cross-platform compatibility. Figure 7 illustrates TICTAC’s workflow  
401 and interfaces, underscoring its defined role in disease–target discovery pipelines. With its focus on  
402 transparency, interpretability, and confidence, TICTAC offers a practical and scientifically rigorous  
403 tool for advancing biomedical research and drug discovery.

404

## 405 **4.4 Validation**

406 The validation process revealed a 10.91% overlap between MedlineGenomics and TICTAC disease-  
407 gene associations for the 193 shared CUIs, indicating some level of consistency while also highlighting  
408 significant differences between the datasets. These discrepancies arise from variations in data  
409 collection methods, terminological standards, and the granularity of gene-disease associations. The low  
410 percentage of overlap underscores the need for improved standardization in disease-gene datasets. This  
411 process emphasized the value of standardized identifiers such as DOIDs and CUIs for aligning and  
412 comparing biomedical data. It also highlighted potential gaps in data completeness or detail, which  
413 could affect the utility of these datasets in downstream research, and reinforced the importance of  
414 integrating multiple sources to achieve a comprehensive understanding of disease-gene relationships.

415

## 416 **5 Conclusion**

417 The conventional "one gene, one function, one trait" paradigm, as critiqued by Visscher et al. (2017),  
418 is no longer adequate for understanding the intricate mechanisms underlying diseases. This shift in  
419 understanding highlights the need for advanced tools that can navigate the multifaceted nature of  
420 genetic and clinical data, enabling researchers to uncover meaningful relationships between genes,  
421 traits, and diseases (Yang et al., 2021). Modern biomedical research demands tools that move beyond  
422 oversimplifications to provide actionable insights grounded in real-world data. TICTAC rises to this  
423 challenge by offering a dynamic platform tailored for drug target hypothesis generation and refinement,  
424 leveraging clinical trial data and metadata to bridge the gap between data complexity and practical  
425 application. Aligned with the increasing focus on interpretable machine learning and explainable  
426 artificial intelligence (XAI) (Adadi & Berrada, 2018), TICTAC employs transparent, evidence-driven  
427 methodologies to elucidate disease–target associations. Its foundation rests solely on clinical trial data  
428 and metadata, supported by rational, intuitive evidence metrics, and underpinned by a robust open-  
429 source pipeline designed for continuous improvement and scalability. TICTAC’s adaptability allows  
430 it to function either as a standalone resource or as a component integrated with other analytical  
431 interfaces. By simplifying the exploration of complex clinical data, TICTAC contributes meaningfully  
432 to the identification and prioritization of drug targets, offering a practical and evolving tool for  
433 advancing biomedical research and translational science.

## 434 **6 Conflict of Interest**

435 *The authors declare that the research was conducted in the absence of any commercial or financial*  
436 *relationships that could be construed as a potential conflict of interest.*

## 437 **7 Author Contributions**

## 438 **8 Funding**

439 9 This work was partially supported by US National Institutes of Health [U24 224370]  
440 ‘Illuminating the Druggable Genome Knowledge Management Center’ (IDG KMC).

## 441 **10 Acknowledgments**

## 442 **11 Reference**

443 Zarin, D. A., & Ide, N. C. (2005). Registering a clinical trial in ClinicalTrials.gov. *Chest*,  
444 128(1), 319–323.

- 445 Tse, T., Fain, K. M., & Zarin, D. A. (2018). How to avoid common problems when using  
446 ClinicalTrials.gov in research: 10 issues to consider. *BMJ*, 361, k1452.
- 447 Laine, C., Horton, R., DeAngelis, C. D., et al. (2007). Clinical trial registration: Looking back  
448 and moving ahead. *New England Journal of Medicine*, 356(26), 2734–2736.
- 449 Riveros, C., Dechartres, A., Perrodeau, E., et al. (2013). Timing and completeness of trial  
450 results posted at ClinicalTrials.gov and published in journals. *PLoS Medicine*, 10(12),  
451 e1001566.
- 452 Prayle, A. P., Hurley, M. N., & Smyth, A. R. (2012). Compliance with mandatory reporting of  
453 clinical trial results on ClinicalTrials.gov: Cross sectional study. *BMJ*, 344, d7373.
- 454 DeVito, N. J., Bacon, S., & Goldacre, B. (2020). Compliance with legal requirement to report  
455 clinical trial results on ClinicalTrials.gov: A cohort study. *Lancet*, 395(10221), 361–369.
- 456 Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., et al. (2016). The FAIR Guiding  
457 Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018.
- 458 Tasneem, A., Aberle, L., Ananth, H., Chakraborty, S., Chiswell, K., McCourt, B. J., &  
459 Pietrobon, R. (2012). The Database for Aggregate Analysis of ClinicalTrials.gov (AACT) and  
460 Subsequent Regrouping by Clinical Specialty. *PLoS ONE*, 7(3), e33677.  
461 <https://doi.org/10.1371/journal.pone.0033677>
- 462 Moffat, J. G., Vincent, F., Lee, J. A., Eder, J., & Prunotto, M. (2017). Opportunities and  
463 challenges in phenotypic drug discovery: an industry perspective. *Nature Reviews Drug  
464 Discovery*, 16(8), 531-543.
- 465 Scannell, J. W., Blanckley, A., Boldon, H., & Warrington, B. (2012). Diagnosing the decline in  
466 pharmaceutical R&D efficiency. *Nature Reviews Drug Discovery*, 11(3), 191-200.
- 467 Hay, M., Thomas, D. W., Craighead, J. L., Economides, C., & Rosenthal, J. (2014). Clinical  
468 development success rates for investigational drugs. *Nature Biotechnology*, 32(1), 40-51.
- 469 Ledford, H. (2011). Drug candidates derailed in case of mistaken identity. *Nature*, 480(7378),  
470 161-162.
- 471 Liu, Q., Liu, Q., Zhang, J., Tao, C., Shen, F., & Wang, Y. (2021). Predicting the clinical  
472 outcomes of phase II/III trials for cancer drugs using the drug's mechanism of action: a  
473 database analysis. *Scientific Reports*, 11(1), 1239.
- 474 Dahlin, J. L., Inglese, J., & Walters, M. A. (2015). Mitigating risk in academic preclinical drug  
475 discovery. *Nature Reviews Drug Discovery*, 14(4), 279-294.
- 476 Muller, P. Y., & Milton, M. N. (2012). The determination and interpretation of the therapeutic  
477 index in drug development. *Nature Reviews Drug Discovery*, 11(10), 751-761.
- 478 Vincent, F., Loria, P., Pregel, M., Stanton, R., Kitching, L., Nocka, K., Doyonnas, R., Stepan,  
479 C., Gilbert, A., Schroeter, T., & Peakman, M. C. (2015). High-throughput screening (HTS) and  
480 multi-parametric data analysis: challenges and promises. *Frontiers in Pharmacology*, 6, 157.

- 481 Huang, R., Southall, N., Wang, Y., Yasgar, A., Shinn, P., Jadhav, A., Nguyen, D. T., & Austin,  
482 C. P. (2011). The NCGC pharmaceutical collection: a comprehensive resource of clinically  
483 approved drugs enabling repurposing and chemical genomics. *Science Translational Medicine*,  
484 3(80), 80ps16.
- 485 Lysenko, I., Kim, H. Y., Emili, A., & Schlessinger, A. (2018). Predicting adverse drug events  
486 by relational learning on clinical notes. *Journal of the American Medical Informatics*  
487 *Association*, 25(3), 246-253.
- 488 Schriml, L. M., Munro, J. B., Schor, M., Olley, D., McCracken, C., Felix, V., Baron, J. A.,  
489 Jackson, R., Bello, S. M., Bearer, C., Lichenstein, R., Bisordi, K., Dialo, N. C., Giglio, M., &  
490 Greene, C. (2022). The Human Disease Ontology 2022 update. *Nucleic Acids Research*,  
491 50(D1), D1255-D1261. <https://doi.org/10.1093/nar/gkab1063>
- 492 McInnes, B. T., Pedersen, T., & Carlis, J. V. (2007). Using UMLS Concept Unique Identifiers  
493 (CUIs) for Word Sense Disambiguation in the Biomedical Domain. *AMIA Annual Symposium*  
494 *Proceedings*, 533-537
- 495 Lowe, D. M., & Sayle, R. A. (2015). LeadMine: A grammar and dictionary driven approach to  
496 entity recognition. *Journal of Cheminformatics*, 7(Suppl 1), S5. [https://doi.org/10.1186/1758-](https://doi.org/10.1186/1758-2946-7-S1-S5)  
497 [2946-7-S1-S5](https://doi.org/10.1186/1758-2946-7-S1-S5)
- 498 Cook, H. V., & Jensen, L. J. (2019). A guide to dictionary-based text mining. In R. Larson & T.  
499 Oprea (Eds.), *Bioinformatics and drug discovery* (pp. 57-75). Humana Press.  
500 [https://doi.org/10.1007/978-1-4939-9089-4\\_5](https://doi.org/10.1007/978-1-4939-9089-4_5)
- 501 Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., Li, Q., Shoemaker, B. A., Thiessen,  
502 P. A., Yu, B., Zaslavsky, L., Zhang, J., & Bolton, E. E. (2021). PubChem in 2021: new data  
503 content and improved web interfaces. *Nucleic Acids Research*, 49(D1), D1388-D1395.  
504 <https://doi.org/10.1093/nar/gkaa971>
- 505 Zdrzil, B., Felix, E., Hunter, F., Manners, E. J., Blackshaw, J., Corbett, S., de Veij, M.,  
506 Ioannidis, H., Lopez, D. M., Mosquera, J. F., Magarinos, M. P., Bosc, N., Arcila, R., Kizilören,  
507 T., Gaulton, A., Bento, A. P., Adasme, M. F., Monecke, P., Landrum, G. A., & Leach, A. R.  
508 (2024). The ChEMBL Database in 2023: a drug discovery platform spanning multiple  
509 bioactivity data types and time periods. *Nucleic acids research*, 52(D1), D1180–D1192.  
510 <https://doi.org/10.1093/nar/gkad1004>
- 511 Sheils, T. K., Mathias, S. L., Kelleher, K. J., Siramshetty, V. B., Nguyen, D.-T., Bologna, C. G.,  
512 Jensen, L. J., Vidović, D., Koleti, A., Schürer, S. C., Waller, A., Yang, J. J., Holmes, J., Bocci,  
513 G., Southall, N., Dharkar, P., Mathé, E., Simeonov, A., & Oprea, T. I. (2021). TCRD and  
514 Pharos 2021: mining the human proteome for disease biology. *Nucleic Acids Research*,  
515 49(D1), D1334–D1346. <https://doi.org/10.1093/nar/gkaa993>
- 516 The UniProt Consortium. (2023). UniProt: the Universal Protein Knowledgebase in 2023.  
517 *Nucleic Acids Research*, 51(D1), D523-D531. <https://doi.org/10.1093/nar/gkac1052>
- 518 U.S. National Library of Medicine. (n.d.). Genetics. MedlinePlus. Retrieved December 11,  
519 2024, from <https://medlineplus.gov/genetics/>



- 520 Oprea, T. I., Bologa, C. G., Brunak, S., Campbell, A., Gan, G. N., Gaulton, A., Gomez, S. M.,  
521 Guha, R., Hersey, A., Holmes, J., Jadhav, A., Jensen, L. J., Johnson, G. L., Karlson, A., Leach,  
522 A. R., Ma'ayan, A., Malovannaya, A., Mani, S., Mathias, S. L., ... Zahoránszky-Köhalmi, G.  
523 (2018). Unexplored therapeutic opportunities in the human genome. *Nature Reviews Drug*  
524 *Discovery*, 17(5), 317-332. <https://doi.org/10.1038/nrd.2018.14>
- 525 Santos, R., Ursu, O., Gaulton, A., Bento, A. P., Donadi, R. S., Bologa, C. G., Karlsson, A., Al-  
526 Lazikani, B., Hersey, A., Oprea, T. I., & Overington, J. P. (2017). A comprehensive map of  
527 molecular drug targets. *Nature reviews. Drug discovery*, 16(1), 19–34.  
528 <https://doi.org/10.1038/nrd.2016.230>
- 529 Lauritsen, K. M., Nielsen, B. R. R., Tolbod, L. P., Johannsen, M., Hansen, J., Hansen, T. K.,  
530 Wiggers, H., Møller, N., Gormsen, L. C., & Søndergaard, E. (2021). SGLT2 inhibition does not  
531 affect myocardial fatty acid oxidation or uptake, but reduces myocardial glucose uptake and  
532 blood flow in individuals with type 2 diabetes: A randomized double-blind, placebo-controlled  
533 crossover trial. *Diabetes*, 70(3), 800-808. <https://doi.org/10.2337/db20-0921>
- 534 Kimura, G. (2016). Importance of inhibiting sodium-glucose cotransporter and its compelling  
535 indication in type 2 diabetes: Pathophysiological hypothesis. *Journal of the American Society*  
536 *of Hypertension: JASH*, 10(3), 271–278. <https://doi.org/10.1016/j.jash.2016.01.009>
- 537 Huang, K., Luo, X., Liao, B., & et al. (2023). Insights into SGLT2 inhibitor treatment of  
538 diabetic cardiomyopathy: Focus on the mechanisms. *Cardiovascular Diabetology*, 22(86).  
539 <https://doi.org/10.1186/s12933-023-01816-5>
- 540 National Institutes of Health. (2024, December 2). Illuminating the Druggable Genome (IDG).  
541 <https://commonfund.nih.gov/idg>
- 542 Nguyen, D. T., Mathias, S., Bologa, C., Brunak, S., Fernandez, N., Gaulton, A., Hersey, A.,  
543 Holmes, J., Jensen, L. J., Karlsson, A., Liu, G., Ma'ayan, A., Mandava, G., Mani, S., Mehta, S.,  
544 Overington, J., Patel, J., Rouillard, A. D., Schürer, S., Sheils, T., Simeonov, A., Sklar, L. A.,  
545 Southall, N., Ursu, O., Vidovic, D., Waller, A., Yang, J., Jadhav, A., & Oprea, T. I. (2017).  
546 Pharos: Collating protein information to shed light on the druggable genome. *Nucleic Acids*  
547 *Research*, 45(D1), D995–D1002. <https://doi.org/10.1093/nar/gkw1072>
- 548 Ghousaini, M., Mountjoy, E., Carmona, M., Peat, G., Schmidt, E. M., Hercules, A., Fumis, L.,  
549 Miranda, A., Carvalho-Silva, D., Buniello, A., Burdett, T., Hayhurst, J., Baker, J., Ferrer, J.,  
550 Gonzalez-Uriarte, A., Jupp, S., Karim, M. A., Wright, D., Suveges, D., & Dunham, I. (2021).  
551 Open Targets Genetics: systematic identification of trait-associated genes using large-scale  
552 genetics and functional genomics. *Nucleic Acids Research*, 49(D1), D1311–D1320.  
553 <https://doi.org/10.1093/nar/gkaa840>
- 554 Ochoa, D., Hercules, A., Carmona, M., Suveges, D., Gonzalez-Uriarte, A., Malangone, C.,  
555 Miranda, A., Fumis, L., Carvalho-Silva, D., Spitzer, M., Peat, G., Baker, J., Ferrer, J., Raies,  
556 A., Wright, D., Imker, H. J., Schwartzentruber, J., Hulcoop, D. G., McDonagh, E. M., &  
557 Dunham, I. (2021). Open Targets Platform: Supporting systematic drug–target identification  
558 and prioritisation. *Nucleic Acids Research*, 49(D1), D1302–D1310.  
559 <https://doi.org/10.1093/nar/gkaa1027>

- 560 Sheils, T. K., Mathias, S. L., Kelleher, K. J., Siramshetty, V. B., Nguyen, D.-T., & Bologna, C.  
561 G. (2021). TCRD and Pharos 2021: Mining the human proteome for disease biology. *Nucleic*  
562 *Acids Research*, 49(D1), D1334–D1346. <https://doi.org/10.1093/nar/gkaa993>
- 563 Yang, J. J., Grissa, D., Lambert, C. G., Bologna, C. G., Mathias, S. L., Waller, A., Wild, D. J.,  
564 Juhl Jensen, L., & Oprea, T. I. (2021). TIGA: target illumination GWAS analytics.  
565 *Bioinformatics*, 37(21), 3865–3873. <https://doi.org/10.1093/bioinformatics/btab427>
- 566 Servick, K. (2020). Pandemic could add noise to clinical trial data. *Science*, 368(6492), 693.  
567 <https://doi.org/10.1126/science.368.6492.693>
- 568 Chen, Y., Lawrence, J., Hung, H. M. J., et al. (2021). Methods for employing information  
569 about uncertainty of ascertainment of events in clinical trials. *Therapeutic Innovation &*  
570 *Regulatory Science*, 55(1), 197–211. <https://doi.org/10.1007/s43441-020-00206-3>
- 571 Verzilli, C. J., & Carpenter, J. R. (2002). Assessing uncertainty about parameter estimates with  
572 incomplete repeated ordinal data. *Statistical Modelling*, 2(3), 203–215.  
573 <https://doi.org/10.1191/1471082x02st033oa>
- 574 Chen, Y., Chen, Y., Lawrence, J., Hung, H.M., & Stockbridge, N. (2020). Methods for  
575 employing information about uncertainty of ascertainment of events in clinical trials.  
576 *Therapeutic Innovation & Regulatory Science*, 55(1), 197–211.  
577 <https://doi.org/10.1007/S43441-020-00206-3>
- 578 Frampton, G., & Shepherd, J. (2008). Ambiguity of study population analysis and reporting in  
579 asthma clinical trials. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen*,  
580 102, 52–52. <https://doi.org/10.1016/J.ZEFQ.2008.06.015>
- 581 Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable  
582 artificial intelligence (XAI). *IEEE Access*, 6, 52138–52160.  
583 <https://doi.org/10.1109/ACCESS.2018.2870052>

## 584 **12 Supplementary Material**

## 585 **13 Data Availability Statement**

586 The datasets generated and analyzed for this study as well as the source code can be found in the  
587 TICTAC repository (<https://github.com/unmtransinfo/TICTAC>).

588

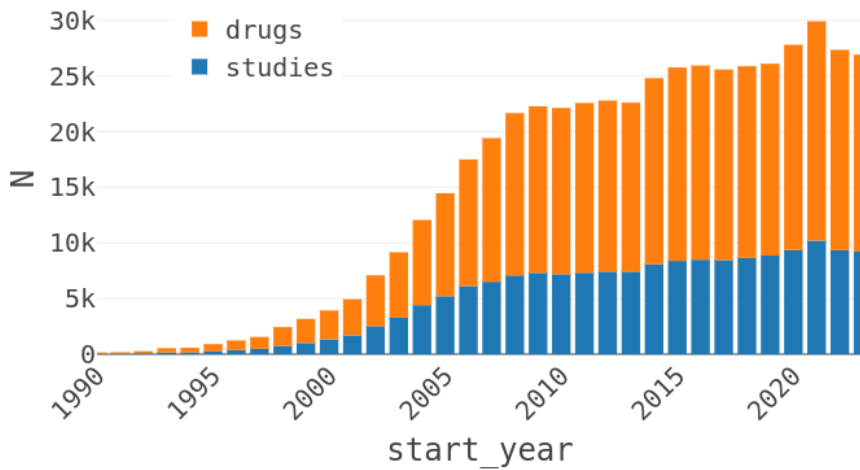
589

590

591

592 Fig. 1. Clinical Trial study counts by year indicating growth and trends

## AACT: Studies and drugs 1990-2023



593

594

595

596

597

598

599

600

601

602

603

604

605

606

607

608

609

610 Fig. 2. TICTAC dashboard. The image displays genes currently associated with the disease 'type 2  
611 diabetes/insulin resistance' (DOID:9352)

**Interactive TICTAC Disease-Target Dashboard**

igDTL Filter:

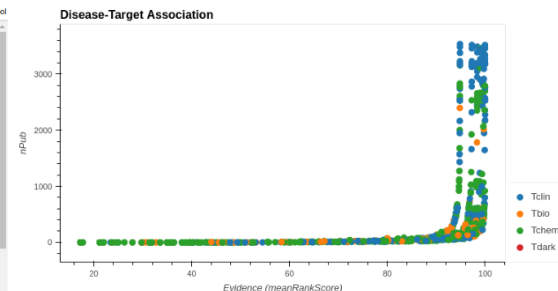
Gene Filter:

Y-axis Variable:

Page Size: 40

Page Number: 2

#	doid_unprot	nDiseases	nDrug	nStud	nPub	nStudyNewnes	nPublicationWt	meanRankScore	disease_term	drug_name	igDTL	gene_symbol
0	DOID:9352_Q7L2E3	1	1	1	0	0.0	NaN	17.25382	type 2 diabetes	Sorafenib	Tchem	DHX30
1	DOID:9352_P12302	1	1	1	1	0.29	0.25	21.76116	type 2 diabetes	Mozobil	Tchem	CXCR5
2	DOID:9352_P41368	31	1	1,160	3,199	1,201.85	1,786.25	94.860756	insulin resistan	Topiramate	Tchem	MCSR
3	DOID:9352_Q9AKV0	7	2	16	96	34.83	60.75	92.489059	type 2 diabetes	N-acetyl cystei	Tclin	HDAC9
4	DOID:9352_P24385	2	2	4	0	0.0	NaN	40.458111	insulin resistan	Genistein	Tchem	CCND1
5	DOID:9352_Q95283	4	1	4	7	3.27	2.75	64.563164	type 2 diabetes	Tadalafil	Tclin	PDE8B
6	DOID:9352_P56524	7	1	16	95	34.26	60.5	89.076843	insulin resistan	Curcumin	Tclin	HDAC4
7	DOID:9352_Q9NSY1	2	6	6	0	0.0	NaN	47.88765	type 2 diabetes	Alpelisab	Tchem	BMP2K
8	DOID:9352_P48736	3	1	7	8	2.42	4.0	66.405647	type 2 diabetes	Alpelisab	Tclin	PIK3CG
9	DOID:9352_Q85XP3	1	2	1	0	0.0	NaN	21.174465	type 2 diabetes	Quercetin	Tchem	DDX42
10	DOID:9352_Q99598	2	1	2	0	0.0	NaN	29.779272	insulin resistan	Eltagic Acid	Tbio	TSNAX
11	DOID:9352_Q8NH87	1	6	1	23	7.1	12.0	70.390663	type 2 diabetes	thymol	Tdark	OR5K1
12	DOID:9352_P09960	1	1	1	0	0.0	NaN	17.25382	Type 2 diabete	Captopril	Tchem	LTA4H
13	DOID:9352_P41595	34	3	1,295	3,489	1,349.29	1,959.0	98.389687	type 2 diabetes	empagliflozin	Tclin	HTR2B
14	DOID:9352_P24941	16	1	192	562	198.05	382.5	94.242134	insulin resistan	Obeticholic aci	Tchem	CDK2



612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630

631 Fig. 3. Examples of understudied (Tbio) genes for disease 'type 2 diabetes/insulin resistance' in  
 632 TICTAC

#	doiid_uniprot	nDiseases	nDrug	nStud	nPub	nStudyNewnes	nPublicationWeigh	meanRankSco	disease_term	drug_name	idgTDL	gene_symbol
0	DOID:9352_Q8NI60	2	5	5	0	0.0	NaN	46.335445	type 2 diabetes	Dasatinib	Tbio	COQ8A
1	DOID:9352_P31152	1	3	2	0	0.0	NaN	31.432911	type 2 diabetes	Sorafenib	Tbio	MAPK4
2	DOID:9352_Q9UBK2	5	1	12	8	2.85	3.75	69.92807	insulin resistan	prednisone	Tbio	PPARGC1A
3	DOID:9352_Q9NRP7	1	4	3	0	0.0	NaN	36.315186	type 2 diabetes	Sorafenib	Tbio	STK36
4	DOID:9352_Q99755	1	4	3	0	0.0	NaN	36.315186	type 2 diabetes	Ruboxistaurin	Tbio	PIP5K1A
5	DOID:9352_P63000	3	2	4	52	15.06	26.25	83.482651	type 2 diabetes	Berberine hydr	Tbio	RAC1
6	DOID:9352_O15438	31	1	912	2,396	908.77	1,332.75	94.828827	type 2 diabetes	Pioglitazone	Tbio	ABCC3
7	DOID:9352_Q9HAW8	8	25	42	109	41.03	54.75	97.87738	type 2 diabetes	Bupropion	Tbio	UGT1A10
8	DOID:9352_Q96CH1	6	1	14	30	13.34	16.0	82.17614	type 2 diabetes	Exendin-4	Tbio	GPR146
9	DOID:9352_P21439	3	2	3	2	0.86	0.5	53.197952	insulin resistan	Digoxin	Tbio	ABCB4
10	DOID:9352_O00762	2	1	2	39	7.38	19.5	72.307777	type 2 diabetes	Candesartan c	Tbio	UBE2C
11	DOID:9352_Q8NEV4	1	3	2	0	0.0	NaN	31.432911	type 2 diabetes	Dasatinib	Tbio	MYO3A
12	DOID:9352_Q01813	2	5	5	0	0.0	NaN	46.335445	type 2 diabetes	Dasatinib	Tbio	PFKP
13	DOID:9352_Q9Y230	1	1	1	0	0.0	NaN	17.25382	type 2 diabetes	Sorafenib	Tbio	RUVBL2
14	DOID:9352_Q14191	3	2	4	14	6.66	9.5	75.701469	type II diabetes	buspirone	Tbio	WRN
15	DOID:9352_P42679	2	1	3	0	0.0	NaN	31.889994	type 2 diabetes	Dasatinib	Tbio	MATK
16	DOID:9352_P23919	1	4	3	5	1.67	3.5	60.488865	insulin resistan	Lamivudine	Tbio	DTYMK
17	DOID:9352_P14859	8	1	33	141	53.55	68.25	91.209229	type 2 diabetes	Pioglitazone	Tbio	POU2F1
18	DOID:9352_Q9H3S4	1	1	1	2	0.37	1.0	29.954967	type II diabetes	Thiamine	Tbio	TPK1
19	DOID:9352_P51843	2	2	7	21	9.32	11.75	78.363693	insulin resistan	Colchicine	Tbio	NR0B1
20	DOID:9352_Q8NA19	1	1	1	0	0.0	NaN	17.25382	type 2 diabetes	oxybutynin	Tbio	L3MBTL4
21	DOID:9352_Q8N5K1	11	5	201	606	202.28	370.25	98.922913	insulin resistan	Pioglitazone	Tbio	CISD2

633

634

635

636

637

638

639

640

641

642

643

644

645

646

647 Fig. 4. Provenance for association between gene MCR4 (Melanocortin receptor 4



648 ) and disease ' type 2 diabetes/insulin resistance'.

<b>TICTAC Provenance Dashboard</b>			
<b>DISEASE: <a href="#">DOID:9352 (Type 2 Diabetes/Insulin resistance)</a></b>			
<b>GENE: <a href="#">MCR4 (Melanocortin receptor 4)</a></b>			
Filter by NCT ID or PMID		Filter by Reference Type	
None		None	
Page Size: 20		Page Number: 1	
nct_id	reference_type	pmid	citation
NCT02796378	BACKGROUND	21424396	Larsen S, Stride N, Hey-Mogensen M, Hansen CN, Andersen JL, Madsbad S, Worm D, Helge JW, Dela F. Increased mitochondrial substrate sensitivity in skeletal muscle of patients with type 2 diabetes. <i>Diabetologia</i> . 2011 Jun;54(6):1427-36. doi: 10.1007/s00125-011-2098-4. Epub 2011 Mar 18.
NCT04841096	BACKGROUND	29405543	Khunti K, Kosiborod M, Ray KK. Legacy benefits of blood glucose, blood pressure and lipid control in individuals with diabetes and cardiovascular disease: Time to overcome multifactorial therapeutic inertia? <i>Diabetes Obes Metab</i> . 2018 Jun;20(6):1337-1341. doi: 10.1111/dom.13243. Epub 2018 Mar 11.
NCT04841096	BACKGROUND	28155046	Bianchi C, Daniele G, Dardano A, Miccoli R, Del Prato S. Early Combination Therapy with Oral Glucose-Lowering Agents in Type 2 Diabetes. <i>Drugs</i> . 2017 Mar;77(3):247-264. doi: 10.1007/s40265-017-0694-4.
NCT05105919	RESULT	20207433	Lordkipanidze M, Pharand C, Schampaert E, Palisaitis DA, Diodati JG. Heterogeneity in platelet cyclooxygenase inhibition by aspirin in coronary artery disease. <i>Int J Cardiol</i> . 2011 Jul 1;150(1):39-44. doi: 10.1016/j.ijcard.2010.02.025. Epub 2010 Mar 7.
NCT02225834	BACKGROUND	20816828	Piermartiri TC, Figueiredo CP, Rial D, Duarte FS, Bezerra SC, Mancini G, de Bem AF, Prediger RD, Tasca CI. Atorvastatin prevents hippocampal cell death, neuroinflammation and oxidative stress following amyloid-beta(1-40) administration in mice: evidence for dissociation between cognitive deficits and neuronal damage. <i>Exp Neurol</i> . 2010 Dec;226(2):274-84. doi: 10.1016/j.expneurol.2010.08.030. Epub 2010 Sep 15. Erratum In: <i>Exp Neurol</i> . 2021 Nov;345:113840. doi: 10.1016/j.expneurol.2021.113840.
NCT02949752	DERIVED	3464644	Gupta B, Chee KS, Neo LQ, Tang C, Hariram J, Tan GC, Verma S, Basu S, Appan DP, Ting CC, Abdin E, Lee J. Effect of aripiprazole as an adjunct to atypical antipsychotics on weight and metabolic profile: a 12-week open-label trial. <i>Ther Adv Psychopharmacol</i> . 2021 Oct 9;11:20451253211046765. doi: 10.1177/20451253211046765. eCollection 2021.

649

650

651

652

653

654

655

656

657

658

659

660

661

662

663

664 Fig. 5. TICTAC dashboard, displaying a plot of genes associated with disease 'lung cancer'  
 665 (DOID:1324)

### Interactive TICTAC Disease-Target Dashboard: DOI:1324

idgTDL Filter  Gene Filter  Y-axis Variable

Page Size: 20  Page Number: 1

#	doi_uniprot	nDiseases	nDrug	nStud	nPub	nStudyNewnes	nPublicationWe	meanRankScor	disease_term	drug_name	idgTDL	gene_symbol
0	DOI:1324_P5	4	16	144	101	45.6	63.25	97.723976	lung cancer	Abemaciclib	Tchem	ADK
1	DOI:1324_P0	2	2	16	52	22.36	28.5	86.472595	lung cancer	5-fluorouracil	Tchem	HSP90AB1
2	DOI:1324_P3	6	170	398	705	277.11	437.25	99.913719	lung cancers	Carboplatin	Tclin	BDKRB2
3	DOI:1324_P2	4	9	76	56	22.01	28.25	94.851616	lung cancer	Dasatinib	Tchem	MAK
4	DOI:1324_Q9	4	45	139	189	81.36	134.0	98.989281	lung cancer	Azacitidine	Tbio	ATXN2
5	DOI:1324_Q8	4	1	79	57	22.58	28.5	87.290588	lung cancers	Erlotinib	Tchem	HIPK1
6	DOI:1324_P0	1	3	6	2	0.84	2.25	54.988091	lung cancer	captopril	Tclin	MME
7	DOI:1324_P2	1	4	7	5	1.57	4.75	64.866219	lung cancer	bortezomib	Tchem	MMP8
8	DOI:1324_Q9	1	2	5	2	0.84	2.25	51.415077	lung cancer	bortezomib	Tbio	USP18
9	DOI:1324_P3	4	9	76	56	22.01	28.25	94.851616	lung cancer	Selumetinib	Tchem	TIE1
10	DOI:1324_P4	4	1	80	57	22.58	28.5	87.297974	lung cancer	BKM120	Tchem	ZAP70
11	DOI:1324_P2	1	3	2	0	0.0	NaN	31.432911	lung cancer	carboplatin	Tclin	RXRB
12	DOI:1324_Q9	1	1	2	3	0.64	1.5	40.592606	lung cancer	Pioglitazone	Tchem	CPT1B
13	DOI:1324_P3	4	16	144	101	45.6	63.25	97.723976	lung cancer	Erlotinib	Tchem	DNAJA1
14	DOI:1324_P1	6	2	465	762	304.16	480.5	96.788269	lung cancers	Dasatinib	Tclin	ADRA2C
15	DOI:1324_Q8	4	1	79	57	22.58	28.5	87.290588	lung cancer	Dasatinib	Tchem	DCLK2
16	DOI:1324_P8	5	1	175	224	93.38	154.75	92.895927	lung cancer	disulfiram	Tchem	SMAD3
17	DOI:1324_P3	6	30	443	718	289.86	462.5	99.858116	lung cancer	bexarotene	Tclin	ADRA1A
18	DOI:1324_Q8	4	38	83	56	22.01	28.25	95.718948	lung cancer	Dasatinib	Tbio	VRK2
19	DOI:1324_Q9	4	9	79	57	22.58	28.5	94.922493	lung cancer	Dasatinib	Tchem	MARK1

Export Data

666  
667  
668  
669  
670  
671  
672  
673  
674  
675

676 Fig. 6. Provenance for association between **PGR** gene (Uniprot: P06401) ' **Progesterone receptor**'  
677 and disease 'lung cancer'

TICTAC Provenance Dashboard					
Filter by Uniprot		Filter by Reference Type			
P06401		None			
Page Size: 20		Page Number: 1			
doid_uniprot	nct_id	reference_type	pmid	citation	uniprot
DOI:1324_P06401	NCT00002822	RESULT	nan	Thatcher N, Qian W, Girling DJ: Ifosfamide, carboplatin and etoposide with mid-cycle vincristine (ICE-V) versus standard chemotherapy (C) in patients with small cell lung cancer (SCLC) and good performance status (PS): results of an MRC randomized trial (LU21). [Abstract] Proceedings of the American Society of Clinical Oncology 22: A-2489, 619, 2003.	P06401
DOI:1324_P06401	NCT04153734	RESULT	28271869	Seymour L, Bogaerts J, Perrone A, Ford R, Schwartz LH, Mandrekar S, Lin NU, Litiere S, Dancy J, Chen A, Hodi FS, Therasse P, Hoekstra OS, Shankar LK, Wolchok JD, Ballinger M, Caramella C, de Vries EGE; RECIST working group. iRECIST: guidelines for response criteria for use in trials testing immunotherapeutics. <i>Lancet Oncol.</i> 2017 Mar;18(3):e143-e152. doi: 10.1016/S1470-2045(17)30074-8. Epub 2017 Mar 2. Erratum In: <i>Lancet Oncol.</i> 2019 May;20(5):e242. doi: 10.1016/S1470-2045(19)30240-2.	P06401
DOI:1324_P06401	NCT00003299	RESULT	15923572	Niell HB, Herndon JE 2nd, Miller AA, Watson DM, Sandler AB, Kelly K, Marks RS, Perry MC, Ansari RH, Otterson G, Ellerton J, Vokes EE, Green MR; Cancer and Leukemia Group. Randomized phase III intergroup trial of etoposide and cisplatin with or without paclitaxel and granulocyte colony-stimulating factor in patients with extensive-stage small-cell lung cancer: Cancer and Leukemia Group B Trial 9732. <i>J Clin Oncol.</i> 2005 Jun 1;23(16):3752-9. doi: 10.1200/JCO.2005.09.071.	P06401
DOI:1324_P06401	NCT00256282	BACKGROUND	21769667	Eroglu Z, Kong KM, Jakowatz JG, Samlowski W, Fruehauf JP. Phase II clinical trial evaluating docetaxel, vinorelbine and GM-CSF in stage IV melanoma. <i>Cancer Chemother Pharmacol.</i> 2011 Oct;68(4):1081-7. doi: 10.1007/s00280-011-1703-z. Epub 2011 Jul 19.	P06401
DOI:1324_P06401	NCT00509366	BACKGROUND	1705771	Potti A, Dressman HK, Bild A, Riedel RF, Chan G, Sayer R, Cragun J, Cottrill H, Kelley MJ, Petersen R, Harpole D, Marks J, Berchuck A, Ginsburg GS, Febbo P, Lancaster J, Nevins JR. Genomic signatures to guide the use of chemotherapeutics. <i>Nat Med.</i> 2006 Nov;12(11):1294-300. doi: 10.1038/nm1491. Epub 2006 Oct 22. Erratum In: <i>Nat Med.</i> 2007 Nov;13(11):1388. <i>Nat Med.</i> 2008 Aug;14(8):889.	P06401

678

679

680

681

682

683

684

685

686

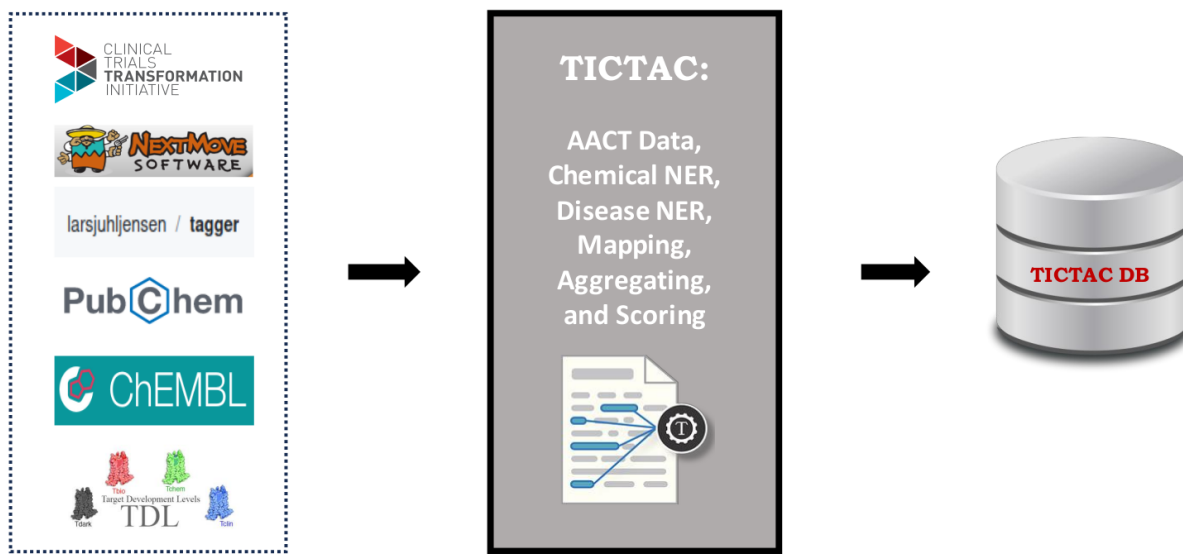
687

688

689

690 Fig 7: TICTAC data sources and interfaces. TICTAC integrates clinical trial data from the AACT db  
 691 and cross referenced other sources to rank disease–target associations. These associations can be  
 692 accessed through the TICTAC github repository.

## TICTAC Workflow



693

694

695

696