Disease-Type-Aware Biological Intelligence Layers for Preventing Symptom Marker False Positives in Computational Drug Target Identification

Kelyn Paul Njeri

NeoForge Labs (Independent Research)

kelyn@neoforgelabs.tech ORCID: 0009-0000-1068-4512

Abstract

Computational drug target identification pipelines that mine biomedical knowledge graphs frequently surface genes that are *statistically* associated with a disease but are biologically inappropriate therapeutic targets. We show that this problem is particularly severe for infectious diseases, where neurological symptom markers (GABA receptors, voltage-gated ion channels, neuronal structural proteins) and inflammatory cytokines (TNF, IL-6) consistently rank among the top causal targets due to strong statistical associations with disease phenotypes. We present a **three-layer biological intelligence system** consisting of (1) a disease-type-aware target classifier with curated blocklists covering 9 gene families (119+ genes) and a 6-step classification pipeline, (2) a tissue expression filter integrating the Human Protein Atlas, and (3) a known-target validator with curated ground truth for 7 diseases. The system correctly reclassifies GABRD as a *symptom marker* in malaria (not a drug target), blocks TNF/IL-6 as *correlational* in infectious contexts while preserving them as valid targets for autoimmune diseases, and identifies host–pathogen invasion receptors (GYPA, CCR5, ACE2) through curated interaction databases and Gene Ontology terms. Validation against FDA-approved drugs across seven diseases achieves a mean $F_1 = 0.474$ (range 0.333–0.700), with zero known false positives promoted to CAUSAL across all diseases. All three layers are disease-context-dependent: the same gene may be classified differently for malaria vs. Alzheimer's disease, reflecting genuine biological differences in target validity. The system is released as part of the open-source NeoRx platform.

Keywords: drug target identification, false positive reduction, disease classification, tissue expression, symptom markers, biological knowledge

1 Introduction

1.1 The False Positive Problem

Modern drug target identification pipelines integrate multiple data sources—genetic associations (GWAS, Monarch Initiative), protein–protein interactions (STRING), pathway databases (KEGG, Reactome), and structural data (PDB)—to construct disease-specific knowledge graphs and identify causal targets via statistical or causal inference methods [1, 2]. While these approaches have demonstrated considerable success, they share a fundamental limitation: **purely statistical methods cannot distinguish between genes that cause disease progression and genes that respond to disease**.

Consider malaria. Mining a knowledge graph for genes causally associated with malaria reliably surfaces GABRD (GABA-A receptor delta subunit), SCN2A (voltage-gated sodium channel Nav1.2), and TNF (tumour necrosis factor alpha) among the top hits. These associations are

real: GABRD is strongly associated with cerebral malaria seizures [3], SCN2A with neurological manifestations [4], and TNF with the inflammatory cascade driving severe malaria [5]. However, none are appropriate antimalarial drug targets:

- **GABRD** is a neuronal receptor. Blocking it would treat seizures (a *symptom* of cerebral malaria), not the underlying Plasmodium infection.
- **SCN2A** modulates neuronal excitability. Its association reflects neurological damage, not parasitic mechanism.
- **TNF** is an inflammatory cytokine elevated in response to parasitaemia. Blocking TNF would suppress the immune response, potentially worsening the infection.

In contrast, the validated antimalarial targets are DHFR/DHPS (parasite enzymes in the folate pathway, targeted by pyrimethamine/sulfadoxine) and host invasion receptors such as GYPA, CR1, BSG, and DARC that the parasite directly binds during red blood cell invasion [6, 7].

1.2 Why Context Matters

The challenge is further complicated by the fact that target validity is *disease-context-dependent*. TNF is a correlational marker in malaria (blocking it would be harmful), but it is a validated therapeutic target in rheumatoid arthritis (infliximab, adalimumab) and Crohn's disease. GABA receptors are symptom markers in infectious diseases but may be legitimate targets in neurodegenerative conditions where GABAergic signalling is directly impaired.

No existing computational pipeline, to our knowledge, incorporates disease-type-aware classification logic to handle these distinctions. Most treat all diseases identically, applying the same scoring and ranking algorithms regardless of whether the disease is infectious, metabolic, neurodegenerative, or autoimmune.

1.3 Contributions

We present a three-layer biological intelligence system that addresses these challenges:

1. **Target Classifier**: A 6-step rule-based classification pipeline with disease-type awareness, curated blocklists covering 9 neurological/systemic gene families, inflammatory cytokine detection, and host–pathogen interaction lookup
2. **Tissue Expression Filter**: A multi-resolution tissue annotation layer integrating the Human Protein Atlas (HPA) with curated fallback data for 63 genes across 12 diseases
3. **Known-Target Validator**: A post-hoc quality assurance layer computing precision, recall, and F against curated ground truth from FDA-approved drugs and WHO Essential Medicines

2 Related Work

2.1 Computational Target Identification

Open Targets [8] integrates genetic associations, transcriptomics, and literature mining to score target–disease associations, but does not distinguish causal mechanisms from correlational ones. The Connectivity Map [9] identifies drug targets through transcriptomic signatures but lacks disease-type-aware filtering. Network-based approaches [10] use guilt-by-association on protein interaction networks, which is susceptible to the same false positive problem we address.

2.2 Gene Expression Filtering

The Human Protein Atlas [11] provides tissue-level expression data for 80% of human protein-coding genes. Several groups have used HPA data to filter drug targets by tissue relevance [12], but none integrate this with disease-type-aware classification logic.

2.3 Target Validation Databases

DrugBank [13], ChEMBL [14], and the Therapeutic Target Database (TTD) [15] provide curated mappings from drugs to targets. These have been used for retrospective validation of target identification methods [16], but not as real-time quality gates within automated pipelines.

3 Methods

3.1 Disease Type Taxonomy

We define a 9-category disease taxonomy that captures the biological distinctions relevant to target classification:
ccc

Category	Code	Example Diseases
Infectious (viral)	`INFECTIOUS_VIRAL`	HIV, Ebola, COVID-19, Hepatitis B/C, Influenza, Dengue, Zika
Infectious (parasitic)	`INFECTIOUS_PARASITIC`	Malaria, Trypanosomiasis, Leishmaniasis, Schistosomiasis
Infectious (bacterial)	`INFECTIOUS_BACTERIAL`	Tuberculosis, Cholera, Meningitis, Pneumonia, Sepsis
Cancer	`CANCER`	Lung cancer, Breast cancer, Colorectal cancer, Melanoma, Glioblastoma
Metabolic	`METABOLIC`	Type 2 diabetes, Obesity, NAFLD
Autoimmune	`AUTOIMMUNE`	Rheumatoid arthritis, Lupus, Multiple sclerosis, Crohn's disease
Neurodegenerative	`NEURODEGENERATIVE`	Alzheimer's, Parkinson's, Huntington's, ALS
Genetic	`GENETIC`	Cystic fibrosis, Sickle cell disease
Other	`OTHER`	Fallback category

Disease resolution proceeds in three stages: (1) exact match against a curated dictionary of 44 disease names, (2) substring matching, and (3) heuristic pattern matching (e.g., the presence of "carcinoma", "sarcoma", or "leukemia" implies cancer).

3.2 Symptom Marker Gene Families

We curate 9 gene families (119+ individual genes) that represent neurophysiological and systemic processes. These genes are frequently associated with disease symptoms but are never appropriate direct targets for infectious diseases:
cccc

Family	Count	Examples	Biological Role
GABA receptors	19	GABRA1–6, GABRB1–3, GABRD, GABRE, GABRG1–3	Inhibitory neurotransmission; associated with seizures in cer
Voltage-gated Na channels	14	SCN1A–SCN11A, SCN1B–SCN4B	Neuronal/cardiac action potentials; associated with neurolog
Voltage-gated K channels	17	KCNA1–5, KCNB1–2, KCNC1–2, KCND1–3, KCNQ1–5	Repolarisation; associated with cardiac/neurological mar
Glutamate receptors	18	GRIN1, GRIN2A–D, GRM1–8, GRIA1–4	Excitatory neurotransmission; associated with excitot
Dopamine receptors	5	DRD1–5	Reward/motor circuits; associated with neuropsychiatric
Serotonin receptors	11	HTR1A–B, HTR2A–C, HTR3A–B, HTR4–7	Mood/GI regulation; associated with psychiatric sym
Coagulation factors	14	F2, F5, F7–F13A1, SERPINC1, SERPIND1, PROC, PROS1, THBD	Haemostasis cascade; associated with DIC in severe in
Acute phase proteins	11	CRP, SAA1–2, HP, HPR, ORM1–2, SERPINA3, FGA, FGB, FGG	Hepatic acute phase response; associated with systemic in
Neuronal structural	10	NEFH, NEFL, NEFM, MAP2, MAPT, TUBB3, SYN1–2, SYP, SNAP25	Axonal/synaptic structure; associated with neurodegenera

For $O(1)$ lookup, all genes are compiled into a flat set, and a reverse mapping (gene $\rightarrow$ family) is maintained for generating human-readable classification reasons.

Critical exception: For neurodegenerative diseases (Alzheimer's, Parkinson's, Huntington's, ALS), symptom marker genes are *not* reclassified. GABA receptors and ion channels may be legitimate targets when GABAergic or voltage-gated channel dysfunction is part of the disease mechanism, not merely a symptom.

3.3 Target Classification Pipeline

Each candidate target is classified through a 6-step pipeline:

3.3.1 Step 1: Symptom Marker Blocklist

$$\text{If gene} \in \mathcal{S}_{\text{markers}} \text{ AND disease is infectious} \implies \text{HOST\_SYMPTOM}$$

where $\mathcal{S}_{\text{markers}}$ is the set of 119+ curated symptom marker genes. The classification includes the family name: *"GABRD is a GABA_receptors family member — these are symptom markers in infectious diseases, not therapeutic targets."*

Neurodegenerative exception: If the disease type is `NEURODEGENERATIVE`, this step is skipped, allowing genes like MAPT (tau) to pass through as valid targets.

3.3.2 Step 2: Inflammatory Cytokine Check

For the specific cytokines TNF, IL6, IL1B, CXCL8, and IL10 in infectious disease contexts:

$$\text{If gene} \in \{\text{TNF, IL6, IL1B, CXCL8, IL10}\} \text{ AND infectious} \implies \text{HOST\_SYMPTOM}$$

Rationale: Cytokine elevation is a *consequence* of immune activation during infection, not a cause of the infection itself. Blocking these cytokines (e.g., anti-TNF therapy in sepsis) risks immunosuppression and has failed in clinical trials [17].

Ordering rationale: The cytokine check is ordered *before* the immune receptor check (Step 3), ensuring that TNF and IL6 are classified as `HOST_SYMPTOM` rather than `HOST_IMMUNE` for infectious diseases. Without this ordering, the immune receptor step's prefix matching would catch these cytokines first and assign a less severe classification.

3.3.3 Step 3: Immune Receptor Detection

Detected by two mechanisms: - **Prefix matching**: Gene symbol starts with TLR, NOD, NLRP, RIG, DDX58, IFIH1, STING, TMEM173, MAVS, or HLA- - **Set membership**: Gene is in a curated set of 30+ immune genes (CD4, CD8A, IFNG, IL-2, STAT1, IRF3, etc.)

$$\text{Classification} = \begin{cases} \text{HOST\_IMMUNE} & \text{if disease is infectious} \\ \text{HOST\_INVASION} & \text{if disease is autoimmune} \end{cases}$$

This reflects the biological reality that immune modulation is *adjunctive* for infections (treating the response, not the cause) but *primary* for autoimmune diseases (where the immune system itself is the problem).

3.3.4 Step 4: Host–Pathogen Interaction Lookup

For infectious diseases, we check whether the gene encodes a known host protein that the pathogen directly binds or hijacks for cell entry:

Curated invasion targets (14 genes):
cc

Pathogen	Host Receptors
P. falciparum (malaria)	GYPA, GYPB, GYPC (glycophorins), CR1, DARC, BSG
HIV	CCR5, CXCR4, CD4
SARS-CoV-2	ACE2, TMPRSS2, NRP1
Ebola	NPC1
General	LAMP1

Additionally, Gene Ontology annotations are checked: - GO:0046718 (viral entry into host cell) - GO:0044409 (entry into host) - GO:0020002 (host cell plasma membrane)

$$\text{If invasion target or GO match} \implies \text{HOST\_INVASION}$$

Host invasion targets are the second-highest priority after pathogen-direct targets, as they represent the physical interface between host and pathogen.

3.3.5 Step 5: Evidence Quality Check

If the target has no source attribution (`source` field missing or empty) and no evidence streams:

$$\text{If source} = \emptyset \text{ AND evidence} = 0 \implies \text{CORRELATIONAL}$$

3.3.6 Step 6: Default Classification

Disease Type	Default Classification
Cancer	`HOST_INVASION` (driver mutation logic)
All others (with evidence)	`HOST_INVASION`

3.4 Tissue Expression Filter

The tissue filter annotates each candidate target with tissue expression data, adding biological context without removing targets.

3.4.1 Tissue Resolution Hierarchy

Given a gene symbol, the primary tissue of expression is resolved via a three-tier hierarchy:

1. **In-memory cache**: $O(1)$ lookup from prior queries
2. **Curated fallback map**: 63 manually curated gene → tissue mappings covering all blocklisted families, key immune genes, and established drug targets
3. **Human Protein Atlas API**: REST query to `https://www.proteinatlas.org/{gene}.json`, parsing RNA tissue-specific nTPM data to find the tissue with maximum expression

3.4.2 Disease–Tissue Mapping

For 12 diseases, we define the biologically relevant tissues:

Disease	Relevant Tissues
Malaria	blood, liver, spleen, bone marrow, placenta
HIV	blood, lymph node, gut, brain, tonsil
Ebola	blood, liver, spleen, lymph node, lung
Tuberculosis	lung, lymph node, blood, bone marrow
COVID-19	lung, blood, heart, kidney, brain
Type 2 diabetes	pancreas, liver, adipose tissue, skeletal muscle
Lung cancer	lung, lymph node, blood
Breast cancer	breast, lymph node, blood
Alzheimer's	brain, cerebral cortex, hippocampus
Parkinson's	brain, substantia nigra, cerebral cortex
Rheumatoid arthritis	synovium, blood, lymph node, joint
Crohn's disease	intestine, colon, blood, lymph node

3.4.3 Relevance Assessment

Tissue relevance is determined via a curated tissue ontology mapping (`_TISSUE_SYNONYMS`) that normalises ~40 HPA tissue names to canonical organ-level forms (e.g., "cerebral cortex" → "brain", "lymphoid tissue" → "lymph node"). Both the gene's expression tissues and the disease's relevant tissues are canonicalised, and relevance is determined by set intersection:

$$\text{relevant}(g, d) = |\mathcal{T}_{\text{canon}}(g) \cap \mathcal{T}_{\text{canon}}(d)| > 0$$

Genes with unknown tissue expression receive the benefit of the doubt: "unknown = pass" — if HPA returns no data, the gene passes by default (we cannot reject what we cannot measure).

Design choice: The tissue filter operates as a **boolean gate** that is **independent** of the causal confidence score. Genes that fail the tissue gate (relevant = False) are automatically demoted to CORRELATIONAL, regardless of their confidence score. The confidence score itself is **never modified** by the tissue filter — it remains a pure measure of causal evidence quality. A separate diagnostic annotation (`tissue_coverage`: fraction of overlapping tissues) is recorded for reporting but never used in scoring. This clean separation ensures that confidence scores remain interpretable and that the tissue gate acts as an independent biological criterion.

3.5 Known-Target Validation

The validator provides post-hoc quality assessment by comparing the pipeline's output against curated ground truth.

3.5.1 Ground Truth Curation

For 7 diseases, we curate known targets (from FDA-approved drugs, WHO Essential Medicines, ChEMBL, DrugBank, TTD) and known false targets:
ccc

Disease	Known Targets	Known False Targets
Malaria	7 (DHFR, DHPS, GYPA, GYPB, CR1, BSG, DARC)	8 (GABRD, GABRA1–2, SCN2A, SCN9A, SCN10A, TNF, IL6)
HIV	4 (CCR5, CXCR4, CD4, POL)	3 (TNF, IL6, CRP)
Type 2 diabetes	6 (GLP1R, SLC5A2, DPP4, PPARG, INSR, INS)	2 (CRP, TNF)
Lung cancer	7 (EGFR, ALK, KRAS, PIK3CA, ERBB2, TP53, BRAF)	0
Breast cancer	6 (ERBB2, ESR1, PIK3CA, BRCA1, BRCA2, CDK4)	0
Alzheimer's	5 (BACE1, APP, PSEN1, MAPT, ACHE)	0
Ebola	2 (NPC1, GP)	2 (TNF, IL6)
Total	**37**	**15**

Each known target entry includes the approved drug, organism, and mechanism of action, enabling detailed diagnostic output.

3.5.2 Validation Metrics

Given the set of targets identified as causal by the pipeline ($\mathcal{C}$), the set of all identified targets ($\mathcal{A}$), the set of known true targets ($\mathcal{T}$), and the set of known false targets ($\mathcal{F}$):

$$\text{True Positives: } TP = \mathcal{C} \cap \mathcal{T}$$

$$\text{Known False Positives: } FP_k = \mathcal{C} \cap \mathcal{F}$$

$$\text{Missed Targets: } FN = \mathcal{T} \setminus \mathcal{A}$$

$$\text{Precision} = \frac{|TP|}{|\mathcal{C}|}, \quad \text{Recall} = \frac{|TP|}{|\mathcal{T}|}, \quad F_1 = \frac{2 \cdot P \cdot R}{P + R}$$

3.5.3 Quality Grading

ccc

Grade	Criterion	Interpretation
A	$F_1 > 0.7$	Excellent — pipeline identifies most known targets with high precision
B	$F_1 > 0.4$	Good — reasonable target identification with some misses
C	$F_1 > 0.2$	Marginal — significant false positives or missed targets
F	$F_1 \leq 0.2$	Poor — pipeline output is unreliable for this disease

3.5.4 Diagnostic Warnings

The validator generates structured warnings:

- **False positive detected**: *"FALSE POSITIVE: GABRD was identified as causal but is a known false target — GABA-A receptor delta, cerebral malaria seizure marker"*
- **Missed target**: *"MISSED: DHFR (targeted by pyrimethamine) is an established target not identified"*
- **Low precision**: *" Low precision (0.25): many identified targets are not in the known-target set"*

4 Results

4.1 Case Study: Malaria

The current NeoRx pipeline integrates ChEMBL pathogen-target evidence with the three biological intelligence layers. For malaria, the ChEMBL pipeline identifies *Plasmodium* enzyme targets (DHFR-TS, PPPK-DHPS) that are invisible to human-only databases, while the classifier demotes host symptom markers:

cccc

Gene	Before Layers	After Layers	Reason
DHFR-TS	CAUSAL (0.896)	PATHOGEN_DIRECT	Pathogen enzyme — ChEMBL confirmed
PPPK-DHPS	CAUSAL (0.825)	PATHOGEN_DIRECT	Pathogen enzyme — ChEMBL confirmed
GABRD	CAUSAL (0.72)	HOST_SYMPTOM	GABA_receptors family — symptom marker
TNF	CAUSAL (0.69)	HOST_SYMPTOM	Inflammatory cytokine in infectious context
SCN2A	CAUSAL (0.65)	HOST_SYMPTOM	Sodium_channels family — symptom marker
GYPA	CAUSAL (0.63)	HOST_INVASION	Known PfEBA-175 receptor
BSG	CORRELATIONAL (0.45)	HOST_INVASION	Known PfRH5 receptor (upgraded)

Validation metrics: $F_1 = 0.333$ (Grade C). The pipeline now correctly identifies pathogen enzymes as top-ranked targets, though recall is limited by the `max_genes` parameter which caps the candidate pool before some known targets can enter.

4.2 Case Study: Alzheimer's Disease

For neurodegenerative diseases, the system correctly preserves targets that would be blocked in infectious contexts:

ccc

Gene	Classification	Reason
MAPT	HOST_INVASION	Neurodegenerative exception — MAPT is in neuronal_structural family but is a validated Alzheimer's target
BACE1	HOST_INVASION	Beta-secretase with strong causal evidence
GABRA1	Not reclassified	Neurodegenerative exception — GABA receptors may be relevant

4.3 Disease-Context Dependence

The same gene receives different classifications depending on disease context:

Gene	Malaria	Rheumatoid Arthritis	Alzheimer's
TNF	HOST_SYMPTOM	HOST_INVASION	CORRELATIONAL
GABRD	HOST_SYMPTOM	HOST_SYMPTOM	Not reclassified
TLR4	HOST_IMMUNE	HOST_INVASION	HOST_IMMUNE
CCR5	HOST_INVASION	CORRELATIONAL	CORRELATIONAL

4.4 Tissue Expression Concordance

Analysis of tissue filter annotations across all 7 validated diseases:

Disease	Targets	Tissue-Relevant	Tissue-Irrelevant	Unknown
Malaria	15	8 (53%)	5 (33%)	2 (13%)
HIV	12	7 (58%)	3 (25%)	2 (17%)
Lung cancer	18	14 (78%)	2 (11%)	2 (11%)
Alzheimer's	10	8 (80%)	1 (10%)	1 (10%)

Tissue-irrelevant targets correlate strongly with symptom marker classification, confirming that tissue expression provides an independent validation signal.

5 Discussion

5.1 The Importance of Disease-Type Awareness

Our results demonstrate that disease-type awareness is not merely an optimisation—it is essential for avoiding scientifically harmful false positives. A pipeline that recommends blocking TNF for Ebola treatment could lead to immunosuppression and worse outcomes. The 9-category taxonomy provides sufficient granularity to capture the major distinctions (infectious vs. autoimmune vs. neurodegenerative) while remaining simple enough for interpretability.

5.2 Blocklist Design Philosophy

The blocklist approach is deliberately conservative: we block only genes where there is overwhelming evidence that they represent symptoms rather than mechanisms. The 9 families were selected because they represent well-characterised neurophysiological (GABA, glutamate, sodium, potassium, dopamine, serotonin receptors), haematological (coagulation factors), inflammatory (acute phase proteins), and structural (neuronal cytoskeleton) processes that are never primary disease mechanisms in infectious diseases.

5.3 Boolean Gate Design

After iterative benchmarking, the tissue filter was implemented as a **boolean gate** rather than a soft scoring modifier. Genes whose canonical expression tissues have zero overlap with the disease's canonical tissues are demoted to CORRELATIONAL regardless of their confidence score. This aggressive strategy is justified by empirical results: annotation-only approaches failed to sufficiently penalise tissue-irrelevant targets, which dominated the top ranks for organ-specific diseases. The boolean gate improved precision for lung cancer, breast cancer, and Alzheimer's disease by removing ubiquitously-expressed inflammation markers. To mitigate false negatives, genes with unknown HPA expression pass by default ("unknown = pass"), and the curated `_TISSUE_SYNONYMS` ontology ensures that nomenclature variations (e.g., "cerebral cortex" "brain") do not create spurious rejections.

5.4 Limitations

1. **Incomplete blocklist coverage**: The 119+ genes cover the major neurological and systemic families but may miss rarer symptom markers. The system can be extended by adding new families to the dictionary.

2. **Curated ground truth limitations**: Validation is only possible for the 7 diseases with curated known targets. For rare diseases, no validation baseline exists.

3. **Rule-based vs. learned**: The classification pipeline uses handcrafted rules rather than learned classifiers. While this ensures interpretability and prevents data leakage, it may miss subtle patterns that a trained model could capture.

4. **Candidate pool size**: The `max_genes` parameter (default 20) limits the initial candidate pool, which can exclude known targets for diseases with large target landscapes. Adaptive pool sizing based on disease complexity is a planned improvement.

5. **ChEMBL symbol quality**: Pathogen gene symbols from ChEMBL `component_synonyms` can be noisy (common English words, multi-word protein names). The current mitigation uses a curated bad-symbol blocklist and minimum-length filters, but edge cases remain.

6 Conclusion

The three biological intelligence layers—target classifier, tissue expression boolean gate, and known-target validator—address a critical gap in computational drug target identification: the inability to distinguish disease mechanisms from disease consequences. Combined with the ChEMBL pathogen-target pipeline and organism-aware relevance scoring, the system achieves a mean $F_1 = 0.474$ across seven diseases (range 0.333–0.700), with particularly strong results for lung cancer (Grade A) and meaningful performance on historically difficult infectious diseases. The layers are modular, interpretable, and extensible, and are released as part of the open-source NeoRx platform.

References

[1] Oprea, T. I. et al. "Drug Repurposing from an Academic Perspective." *Drug Discov. Today Ther. Strateg.* 8, 61–69 (2011).

[2] Whirl-Carrillo, M. et al. "Pharmacogenomics Knowledge for Personalized Medicine." *Clin. Pharmacol. Ther.* 92, 414–417 (2012).

[3] Idro, R. et al. "Cerebral Malaria: Mechanisms of Brain Injury and Strategies for Improved Neurocognitive Outcome." *Pediatr. Res.* 68, 267–274 (2010).

[4] Bhatt, S. et al. "The Effect of Malaria Control on Plasmodium falciparum in Africa between 2000 and 2015." *Nature* 526, 207–211 (2015).

[5] Clark, I. A. et al. "The Role of TNF in Cerebral Malaria." *Ann. Trop. Med. Parasitol.* 99, 757–768 (2005).

[6] Cowman, A. F. et al. "Malaria: Biology and Disease." *Cell* 167, 610–624 (2016).

[7] Crosnier, C. et al. "Basigin is a Receptor Essential for Erythrocyte Invasion by Plasmodium falciparum." *Nature* 480, 534–537 (2011).

[8] Ochoa, D. et al. "Open Targets Platform: Supporting Systematic Drug-Target Identification and Prioritisation." *Nucleic Acids Res.* 49, D1302–D1310 (2021).

[9] Subramanian, A. et al. "A Next Generation Connectivity Map." *Cell* 171, 1437–1452 (2017).

[10] Guney, E. et al. "Network-Based In Silico Drug Efficacy Screening." *Nat. Commun.* 7, 10331 (2016).

[11] Uhlén, M. et al. "Tissue-Based Map of the Human Proteome." *Science* 347, 1260419 (2015).

[12] Santos, R. et al. "A Comprehensive Map of Molecular Drug Targets." *Nat. Rev. Drug Discov.* 16, 19–34 (2017).

[13] Wishart, D. S. et al. "DrugBank 5.0." *Nucleic Acids Res.* 46, D1074–D1082 (2018).

[14] Mendez, D. et al. "ChEMBL: Towards Direct Deposition of Bioassay Data." *Nucleic Acids Res.* 47, D930–D940 (2019).

[15] Wang, Y. et al. "Therapeutic Target Database 2020." *Nucleic Acids Res.* 48, D1063–D1073 (2020).

[16] Brown, A. S. & Patel, C. J. "A Standard Database for Drug Repositioning." *Sci. Data* 4, 170029 (2017).

[17] Fisher, C. J. et al. "Treatment of Septic Shock with the Tumor Necrosis Factor Receptor:Fc Fusion Protein." *N. Engl. J. Med.* 334, 1697–1702 (1996).

Funding

This research received no external funding.

Declaration of Interest

The author declares no competing interests. This is independent research conducted under NeoForge Labs.

Data Availability

All source code, classifier implementations, and validation scripts are available at https://github.com/cod3smith/n under a non-commercial source-available licence (free for personal, academic, and research use; commercial use requires a separate licence). The Human Protein Atlas data is publicly available at https://www.proteinatlas.org/.