Causal Human Biology Is All You Need. But How Much Do We Know?

Part 1: What human genetics of lipids and psychiatric disease teach us about the answer to: *What fraction of causal human biology do we actually know?*

I have been asking a version of this question for months now – at lunch, over dinner, on walks with friends and colleagues. For fun, I have gotten into the habit of politely pressuring people for a single number. *Come on, just give me your best guess!*

Almost without exception, as we worked through the answers together, we did not land on a single number. Instead, we debated a structure. This blog is an attempt to write that structure down.

Every one of those conversations eventually ran into the same unexamined assumption behind most discussions of AI and drug discovery: that we already understand causal human biology well enough for a sufficiently capable model to find the cures hiding inside it.

If that assumption is right, as many with computational or technology backgrounds seem to believe, the job left for AI is largely one of connection and interpretation. Compress the literature, integrate the omics, surface the pattern a human missed.

If it is wrong, as many with biological or medical backgrounds seem to believe, if what we are actually missing is not connective tissue between known facts but the causal facts themselves, then AI's job is a different and much harder one. Either it needs to discover genuinely new causal biology through some kind of emergent reasoning, or it needs to be tightly coupled to physical experiment, cycling through hypothesis and test the way a scientist does, just faster.

Which of those is closer to the truth depends on a structure, not a single fact. Working that structure out, then pressure-testing it against real evidence, is what the rest of this piece does.

It’s enough material that I'm splitting this exploration into two parts. This post lays out the framework and tests it against two disease areas: lipid metabolism and psychiatric disease. Part 2 picks up with a single dyslipidemia gene, *FNIP1*, which complicates the whole picture, and I use it to make the argument that when it comes to AI, the missing ingredient is not more computing power or bigger models, but casual human biology itself.

Why a single answer is the wrong question

Although it’s a fun conversation starter, I understand that asking for a single percentage of how much casual human biology is known is an inherently flawed question. The problem is multi-fold: there is no bounded denominator to divide against; the answer varies by disease area and by the level of detail used to define "causal"; and the answer varies by application (drug discovery, for instance, is a different question than disease taxonomy).

What is far more useful, and what spans across disease areas and practical applications, is that the causal question resolves into three tiers, **defined not by what we would like to know but by where the knowledge currently sits today.**

> Call the first **known**: causal findings already established and reported in the published literature today. Call the second **hidden**: knowledge that already exists, right now, somewhere inside data we have already generated, whether genomic, clinical, or experimental, but that we have not yet extracted, because the datasets have not been integrated, the right analysis has not been run, or we simply have not had the computational power to find it. And call the third **unknown**: knowledge that does not exist in any form yet, because the data needed to reveal it has not been generated and the experiments needed to produce that data have not been run. In an older vocabulary, these are roughly known-knowns, known-unknowns waiting to be found, and true unknown-unknowns requiring investigation.

My conversations on these topics also forced me to anchor the question to a single, concrete application, rather than remain abstract. The application I chose, not surprisingly, is drug discovery and development. Drug development is the ultimate referee of causal human biology. The clinical trial is the one place a causal claim gets tested against reality, not against a model, an assay, or a mouse. A trial's verdict is narrower than it looks, though: it tests a conjunction of target, modality, dose, duration, population and endpoint, so a failure tells you something broke without telling you which part.

The causal question is therefore refined to: *how much causal human biology do we know for a given disease area and target to support drug discovery and development?*

Human genetics turns out to be one of the best players on the field for figuring out where a given piece of disease biology sits among the three tiers before the drug discovery and development referee ever makes its call, and a handful of recent published genetic studies let us test this structure directly. I will use several of them, along with other recent publications or commentaries, as running examples throughout this piece [1][2][3][4][5][6][7].

Why does any of this matter? Because AI's job changes with the tier, and the three tiers call for three different jobs.

> Where a piece of biology is **known**, already established and published, AI's job is mostly connection and interpretation: compress the literature, integrate the omics, surface the pattern a human missed.

> Where it is **hidden**, latent somewhere in data we have already generated but not yet extracted, AI's job is to extract it: better models, more compute, and better integration across existing datasets, aimed squarely at surfacing what is already there.

> Where it is **unknown**, data that does not exist yet in any form, AI's job is different again, and harder: facilitating new discovery and experimentation, generating hypotheses and then helping design and interpret the physical experiments needed to produce the data that does not currently exist, cycling through that loop the way a scientist does, just faster.

> *(It is tempting to imagine an AlphaGo "Move 37" here, some emergent capability to infer causal structure with no new data at all. But Move 37 was hidden, not unknown: Go has a complete, cheap, perfect simulator, so its ground truth can be manufactured on demand, and the move was already latent in a fully specified game. Human biology has no such oracle – at least today – and its ground truth is capped by what has actually been measured, and the realistic version of this third job, the one this piece bets on, runs through the experiment, not around it.)*

Matching the right job to the right tier, for a given question about a given piece of biology, is what both Parts 1 and 2 of this essay attempt to work out. And the answers guide us where to invest in AI to make sense of causal human biology.

The empirical signature: why drugs fail

Since the refined causal question is focused on drug discovery and development, the clearest real-world evidence for this gap is also the most consequential one for anyone building or funding drug programs: the clinical trial. It is the one place a causal claim about human biology gets tested against reality. And it fails, overwhelmingly.

> *"More than 90% of the compounds that enter clinical trials fail to demonstrate sufficient safety and efficacy to gain regulatory approval, most of this failure due to the limited predictive value of preclinical models of disease, and our continued ignorance regarding the consequences of perturbing specific targets over long periods of time in humans."*

That framing is from a paper I wrote with David Altshuler and Ed Scolnick in 2013, on validating targets through human genetics [8], and unfortunately, it has held up remarkably well. A 2026 field-wide critique of AI in drug discovery reinforces our original message: the single largest reason for failure is a lack of clinical efficacy, ahead of toxicity and pharmacokinetics combined, meaning the biological mechanism itself was wrong, not the chemistry built around it [1]. Daphne Koller, CEO of insitro, put the same point more bluntly: in most cases, the molecule was engineered just fine, but the mechanism it targeted was wrong [3].

Read as a hit rate on causal hypotheses rather than a generic failure statistic, this suggests our success rate at correctly identifying a causal, clinically actionable disease mechanism sits somewhere on the order of ~10% per attempt, higher for some indications like rare genetic diseases and lower for neurological diseases such as schizophrenia or Alzheimer's. And that 10% is counting only mechanisms good enough to have survived preclinical filtering and reach a human trial in the first place, already a heavily pre-selected, optimistic sample of everything we believe about disease causality.

> **Lesson #1: the answer to how much causal human biology we know is dependent on the application, which for this blog will be drug discovery and development.** My starting point, offered as a proxy to be tested and refined by the rest of this piece rather than a measured fact, is this: we know roughly 10% of causal human biology as it relates to the disease-specific practical application of drug discovery and development.

Human genetics as our best proxy for "truth"

While there are many tools that can be used to evaluate causal human biology, I will focus on one: human genetics. If there is one tool that reliably beats that ~10% clinical trial success rate, it is human genetics, and the reason is based on its fundamental features. A genetic variant is a naturally randomized perturbation, assigned at conception, running for a lifetime, entirely uncorrelated in expectation with the confounders that plague almost every other form of biological evidence. It is, in a real sense, the closest thing biology offers to a controlled experiment that nature runs on our behalf.

I made a version of this argument in a piece written before ChatGPT's wide release, back in early 2022, framed around the idea of "truth" as the scarce ingredient AI needs and mostly lacks in drug discovery: massive, curated datasets of verified positive and negative outcomes, of the kind that let AlphaGo train against millions of self-played games with an unambiguous win/loss signal. Biology rarely offers anything so clean, except, I argued, in human genetics, where an association robust enough to survive rigorous statistical scrutiny is about as close to ground-truth causality as biology gets outside a clinical trial [9].

The empirical case has only strengthened since our original 2013 review. Drug targets with human genetic support are roughly 2.6 times more likely to succeed in the clinic than those without, according to a 2024 refinement of the original Nelson et al. analysis [10][11]. Tellingly, that advantage improves specifically with confidence in the causal gene, not with effect size, allele frequency, or how recently the variant was discovered. Genetics works precisely when it delivers genuine causal resolution, and works less well when the causal gene, cell type, or direction of effect is still ambiguous, which describes most GWAS loci, the large majority of which sit in noncoding DNA and require substantial additional work to resolve.

A caveat on the known tier is worth stating plainly. “Known” here means established and reported, which is not the same as true. Ioannidis’s argument that a large share of published findings are false turns on low prior odds, modest statistical power and bias, and biology is not exempt [12]. Truth runs at right angles to the three tiers: known, hidden and unknown describe where a finding sits, while truth describes whether it holds. A false known is therefore a defect inside the first tier, and it matters here because a model trained on the literature inherits the literature’s error rate. Human genetics is unusually well defended on this axis, because its hypothesis space can be enumerated before the experiment is run and corrected for accordingly. That is also why genetics raises the prior on the experiments Part 2 argues we still have to run.

Testing the framework with human genetics

This leads us to the question of causal human biology through the lens of human genetics across different diseases. What is convenient about human genetics is that the logic is testable across diseases using a shared set of assumptions.

If most of the causal human biology relevant to a given disease is already known, a new, well-powered genetic study of that disease should mostly rediscover it: the genes it turns up should already be familiar, the mechanisms should map onto pathways we already understand, and its strongest signals should converge on the targets of medicines that are already approved. If, on the other hand, only a small fraction of the relevant causal biology is known, the same kind of study should surface unfamiliar genes, implicate tissues or cell types nobody had previously connected to the disease, and reveal patterns of human physiology that do not fit any existing model, and most of the field's approved drugs should fail to map onto whatever the genetics turns up, because those drugs were found before the relevant biology was known, by other routes entirely.

Indeed, I ran a version of this same test once before, back in 2013, on rheumatoid arthritis, recovering most of the field's approved drug targets from genetics alone, though at the time I did not yet have the language to frame it as a question about how much causal human biology we actually know [13].

This logic rests on one assumption worth flagging: that germline genetics is a powerful instrument for discovering novel drug targets for the disease in question. Not every disease is equally amenable to genetic-driven drug discovery. But where the instrument does work, it gives us a direct, empirical way to ask how much causal biology is already banked versus how much remains to be found, rather than simply asserting an answer.

Two recent, comparably large genetic studies let us run exactly this test side by side: one in lipid metabolism and one in neuropsychiatric disease, both drawing on cohorts of roughly a million people. The two returned strikingly different answers, with caveats that I will describe later.

What ~1 million exomes found in the liver

In 2026, a Regeneron Genetics Center team ran an exome-wide association study across just over a million people, using the triglyceride-to-HDL ratio, a simple, mechanistically well-understood biomarker of fat-storage capacity, as the discovery phenotype [5]. Rare coding variants in 59 genes reached exome-wide significance, and a tissue-enrichment analysis pointed overwhelmingly at two systems, liver and adipose tissue, each many-fold enriched over background [5].

Of those 59 genes, the authors report that 44 – fully 75% - were novel rare-coding-variant associations relative to the previous largest analysis of this trait in about 170,000 people [5]. Read literally, three-quarters of the signal was new. In that case, surely this isn't just rediscovering known biology, right?

Not so fast. Novelty of association is not the same as novelty of biology, and for drug discovery it is biology that matters. A gene can be a brand-new statistical association and still land squarely inside territory a lipid geneticist had already mapped. Run that filter and the picture inverts.

Of the 59 genes, seven sit in the LPL pathway, five are apolipoproteins, and 11 are established monogenic lipid-disorder genes, the kind of genes a lipid geneticist would have named before the study ever ran [5]. More directly to the point, 31 of the 59 are already known drug targets, and 23 of those already have an approved medicine or a compound in human clinical trials [5]. What was new was mostly a set of previously unlogged coordinates on a map that already existed, not a redrawing of the territory. Strip out the recognizable positive controls and only about a fifth (or 20%) of the original 59 genes remain as genuinely unanticipated biology, including one, *FNIP1*, with an effect size well beyond the rest [14].

That is the headline, with two honest caveats: many of the 59 signals had no strong prior causal link to cardiometabolic disease, and several of the 'known' ones had themselves first surfaced only in earlier large-scale sequencing of triglycerides and other lipids. The balance still holds – a large, well-powered genetic study of lipid metabolism mostly reinforced a map that was already drawn, leaving only a small, identifiable residue of genuinely new biology. (Remember *FNIP1*. We will come back to it in Part 2.)

What ~1 million cases found in psychiatric disease

At almost exactly the same scale, just over one million cases, the Psychiatric Genomics Consortium's third cross-disorder analysis asked a similar question not of a quantitative trait but of a collection of 14 psychiatric disorders [6]. Five latent genomic factors, not genes or discrete biological nodes, explained roughly 66% of the shared genetic variance across the fourteen conditions and were associated with 238 pleiotropic loci, with schizophrenia and bipolar disorder loading so heavily onto one shared factor that the two disorders are barely separable at the level of individual genetic variants [6].

Run the same test as in dyslipidemias and the answer lands on the opposite side. There is no equivalent, in this paper, of "31 of these are already known drug targets." The resolution stops well short of individual actionable genes. It resolves to broad biological categories, transcriptional regulation for the signal shared across all fourteen disorders, and, for the two best-powered factors, to cell types rather than genes: excitatory neurons for

schizophrenia/bipolar, oligodendrocytes and other glial cells for the shared internalizing (depression/PTSD/anxiety) signal [6]. Indeed, the paper's own translational claim is prospective rather than retrospective. The authors suggest their findings "*could build ... to identify new or repurposed therapeutics*," not that the loci they found already sit under medicines in use [6]. Even the single most pleiotropic locus in the study, a hotspot associated with eight of the fourteen disorders, illustrates the point rather than undermining it: it contains DRD2, the target of essentially every antipsychotic in clinical use, a drug-target relationship established through classical pharmacology in the 1950s, decades before this genetic data existed, and confirmed by the genetics rather than discovered by it [6]. A parallel case sits close by: CHRM4, which encodes the M4 muscarinic receptor, is a leading candidate causal gene at a well-established schizophrenia GWAS locus [15] – genetic support that arrived after pharmacology had already nominated the M4 mechanism, the same confirm-rather-than-discover pattern as DRD2.

This low resolution is not an artifact of underpowering, and two additional lines of evidence make that clear.

First, the paper's own power calculations suggest the current resolution is nowhere near its ceiling. Based on the MiXeR power curves, schizophrenia, the best-powered disorder in the study, would need many millions more effective cases to saturate genome-wide discovery, and major depression, already past a million cases, would need an order of magnitude more still [6].

Second, and more fundamentally, the ceiling itself is low for reasons that predate this study. A 2023 analysis of nearly 400,000 UK Biobank exomes found that the burden of rare coding variants explains only about 1.3% of phenotypic variance on average across 22 traits, and about 2% specifically for schizophrenia and bipolar disorder; common and rare variants converge on the same cell types but the rare-variant signal is concentrated in a tractable number of large-effect, constrained genes [16]. In other words, lipids resolved so cleanly in large part because the trait being mapped was a quantitative biomarker with an already-established causal link to cardiovascular outcomes; psychiatric disease offers no comparable quantitative readout, so the same rare-variant instrument has far less to work with in the first place.

That said, it is important to look at both perspectives, and it would be unfair to claim that genetics plays no role here. A 2024 whole-exome study of depression in 320,000 UK Biobank participants found that the burden of rare damaging coding variants in loss-of-function-intolerant genes is significantly associated with depression risk, with overlapping rare-variant components shared with autism, schizophrenia, and developmental disorders [17]. The signal is real, replicated, and biologically meaningful. It is not, however, resolved to individual, actionable genes the way lipid biology is.

That is the entire point: ***a genuine causal signal can exist while its resolution to a drug target remains a different order of achievement entirely***.

Put the two studies side by side and the contrast is exact rather than impressionistic. At comparable sample size and comparable methodological sophistication, the lipid study's genetic signal collapsed mostly onto genes we already knew, a large share of them already drugged. The psychiatric study's genetic signal collapsed onto five broad liability dimensions and a handful of cell types, with essentially none of it mapping onto existing medicines – a real, replicated, biologically meaningful result, but a different order of resolution entirely. Lipid metabolism looks like a domain where a lot of the relevant causal biology is already known. Psychiatric disease looks like one where most of it is not.

One caveat before this contrast hardens into a slogan. “Lipids” is not the same as “cardiovascular disease.” Dyslipidemia is the tractable exception inside a much larger domain: LDL-receptor biology, PCSK9, and ANGPTL3 are as cleanly resolved as almost anything in human genetics, but that resolution does not transfer to

heart failure, atherosclerotic plaque biology, or arrhythmia, where the causal picture is far closer to the sub-10% end of the scale. Reading an entire domain by its most tractable sub-phenotype overstates how much of it we actually know.

And Lp(a) is well validated genetic target in dyslipidemia, but recent clinical trial data suggest that targeting it may not translate into cardiovascular risk reduction. More data are required to understand why. A failed trial of one agent against a target is evidence about that agent, its dose and its population, and does not by itself retire the underlying genetics.

The same caution runs the other way. “Psychiatric disease” is a symptom-defined clump of conditions, and schizophrenia on its own is a fairer unit of analysis than the fourteen-disorder aggregate; some individual pathways are better resolved than the pooled signal suggests. The unit of analysis, whether a whole domain, a single condition, or a single pathway, moves the estimate as much as the biology does.

> Nevertheless, **for dyslipidemias,** I estimate based on human genetics that the evidence above puts the answer somewhere in the greater-than-20% category: enough of the causal landscape is already mapped that a drug program built on this knowledge should see a real, measurable lift in its probability of success.

> **For neuropsychiatric disease**, I estimate that the same genetic evidence puts the answer in the less-than-10% category – and probably even less than 1%. Most of the causal landscape remains uncharted, and a program here is making a much larger bet on biology nobody has resolved yet.

There are three cautions, however. First, they are read almost entirely through germline genetics, so a low estimate can reflect either biology that genuinely is not yet known or biology that this one instrument – tested only across part of the allele-frequency spectrum – has not yet resolved. It will read differently once rare-variant, somatic, and longitudinal data are brought to bear; neuropsychiatric disease is the cautionary case, where heritability is high, yet the architecture is so diffuse that the resolved fraction stays low. Second, these are estimates, but comparable human genetic studies across these two diseases support the claim. Third, target quality is not a fixed property of the biology: because these estimates are read through a germline instrument, they can undercount targets whose therapeutic value emerges only when they are paired with a sufficiently powerful intervention – a reminder that matching the modality to the mechanism is itself part of the discipline, a principle I develop in my broader R&D framework [18].

> **Lesson #2: the answer to how much causal human biology we know is not one number; it varies sharply by area of biology, by how cleanly the disease phenotype can be defined and measured, and by the source and resolution of the data used to read it, with real consequences for where and how to invest in drug R&D.**

Three tiers of causal human biology

Those two estimates raise a further question about causality, and Alzheimer's disease is the sharpest case. Its causal genes – APP, PSEN1/2, APOE – are about as well characterized as any disease genes in medicine, with unambiguous causal roles and allelic series across familial mutations of varying severity, yet decades of drug development have produced only modest clinical benefit and very few approved products. How can a disease area

have such strong causal evidence, implicated through human genetics, yet still leave it unclear whether the missing piece for drug discovery is hidden or genuinely unknown? Answering this paradox means returning to the known/hidden/unknown structure from the start of this piece and being disciplined about what it applies to for drug discovery.

Here is the key feature that keeps the framework coherent. Known, hidden, and unknown are tiers of causal human biology, and which tier applies depends on the question being asked. The answer to "is a gene causal" is different from the answer to "how do I perturb the gene's protein product in a specific tissue, with a magnitude of effect that is safe and effective for people." The answer might be “known” for the former, but “hidden” or “unknown” for the latter.

Let us explore this further, using human genetics as the source of causal human biology and drug discovery as the application.

> **The causal disease-level question: is this gene or pathway actually causing the phenotype, and in which direction?** This is the question a genetic association, done well, can answer with something close to certainty. A documented allelic series, a graded relationship between how much function a gene has lost or gained and how severe the resulting phenotype is, functions as what David, Ed, and I called an "experiment of nature" in 2013 [8]: about as close to a randomized controlled trial as biology offers outside an actual clinical trial. Modern human genetics, at biobank scale, has gotten remarkably good at moving the causal disease-level question into the "known" tier for a fast-growing list of genes.

> **The causal mechanistic-driver question: through what pathway, and in which cell type or tissue, does the causal effect actually operate?** Knowing that a gene is causal does not tell you how, or where. This question needs a different kind of evidence: functional genomics, perturbation experiments, primary human cells, ideally in the causally relevant tissue. And it is gated by how accessible the causally relevant tissue is – the variable that separated the lipid and psychiatric cases above, where the genetics resolved to liver and adipose in one and to broad neuronal and glial categories in the other. For an accessible tissue like liver or blood, the answer is often already known or, at worst, hidden in data we can readily generate. For an inaccessible one like brain, the same question frequently sits in the unknown tier even after the disease-level question is settled and moving it into "known" requires new experiments rather than better analysis of data we already hold. This is where most of a domain's residual uncertainty tends to hide.

> **The causal therapeutic-intervention question: can this mechanism be safely and durably modulated, across the intended patient population, to produce net clinical benefit?** This is the question a clinical trial verifies and nothing else fully substitutes for. Its answer can sit in "unknown" even when the disease-level and mechanistic-driver questions are both firmly in "known," because modality, target node, dose, safety, delivery, redundancy, and population heterogeneity are questions of their own, not free corollaries of knowing the causal mechanism.

The therapeutic-intervention question is also the one tier where medicine has, in effect, been running a vast if poorly controlled experiment for decades, and the real-world record speaks to it more directly than any other source. Lithium calmed agitated patients some eighty years ago, long before anyone understood why (Cade, 1949); aspirin's cardiovascular and colorectal-prevention signals surfaced in observational cohorts before randomized trials confirmed them; metformin's associations with lower cardiovascular and cancer risk emerged from registries and claims. Genetics tells you a target is causal; exposure data tells you a drug helped a population, which is closer to

what a clinical trial actually adjudicates. So part of what looks “unknown” on this tier may in fact be hidden, latent in prescribing records, registries, claims, and electronic health records we have never read as experiments, which tilts the work toward extraction and compute more than a genetics-only reading suggests.

The crucial point, easy to miss if you talk about "how well we understand gene X" as though that were a single fact: the same gene can have its disease-level question sitting in "known" and its therapeutic-intervention question sitting squarely in "unknown," at the same time. Treating a "known" answer to the disease-level question as though it settled the other two is, I think, the single most common way this literature ends up overstating how much we actually know. This is exactly what FNIP1 illustrates, and it is worth working through in detail.

That’s where we’ll pick up Part 2 of this essay. Sign up to my Substack account and be the first to know when Part 2 is ready!

Resources for further reading

Each entry is labeled with the citation number it corresponds to in the text above, and is listed in the order it first appears.

[1] Bender, A. et al. “Artificial intelligence in drug discovery, what it is, where we stand and the path forward.” Nature Reviews Drug Discovery (2026). The most comprehensive recent critique of AI's translational impact; the source of the proxy-data figures (liver-organoid DILI, thermal-shift binding) and the reductionism/emergence argument via Lazebnik and Anderson.

[2] Dupire, L., Khan, A. A., Karaletsos, T., Kelley, S., Lundberg, E., Ma, J., Paull, E., Quake, S. R., Rabadan, R., Rowan, C., Sims, P., Tavazoie, S., Tsang, J. S., Zhang, M. & Califano, A. “Fifteen challenges for generative AI applications to cell biology.” Cell 189 (September 17, 2026). doi:10.1016/j.cell.2026.07.004. A Leading Edge Perspective proposing fifteen disease-agnostic grand challenges for generative AI in cell biology, most of which still lack effective predictive methodologies, and arguing that insufficient measurement data and multicellular complexity, not compute, are the binding constraints.

[3] Koller, D. “Drug Discovery Has No Magic Wands.” a16z (August 2026). A founder's-eye view of the same argument, organized around disease-to-mechanism/mechanism-to-drug/drug-to-patient, evolutionary conservation as a predictor of tractability, and the feedback-loop-speed critique of agentic AI in biology.

[4] Akinsanya, K. “Studio Ghibli, The Movies & Drug Discovery?” The Human Algorithm (LinkedIn newsletter, August 2026). Source of the “Herd's Law” argument cited in the conclusion: cheaper computational leverage tends to concentrate investment on already-validated targets rather than expand into unexplored biology.

[5] Hindy, G., Adam, M. et al. “FNIP1 variants are associated with favourable metabolism in 1 million humans.” Nature (2026).

[6] Grotzinger, A. D., Werme, J., Peyrot, W. J., Frei, O. et al. “Mapping the genetic landscape across 14 psychiatric disorders.” Nature 649, 406 to 415 (2026).

[7] Ota, M. et al. “Causal modelling of gene effects from regulators to programs to traits.” Nature 650, 399 to 408 (2025). The current state of the art in combining human loss-of-function data with Perturb-seq to build genuine causal graphs: narrow, well-instrumented, and explicit about the limits of scaling further.

[8] Plenge, R. M., Scolnick, E. M. & Altshuler, D. “Validating therapeutic targets through human genetics.” Nature Reviews Drug Discovery 12, 581 to 594 (2013). The foundational framing of genotype-phenotype dose-response curves and “experiments of nature,” worked through the rheumatoid arthritis example.

[9] Plenge, R. M. “What is truth?” plengegen.com (2022). Written before the current wave of generative AI, this piece first made the case that “truth,” verified, causally-anchored ground truth, rather than data volume, is the binding constraint on AI's usefulness in drug discovery, and proposed human genetics as the best available source of it.

[10] Nelson, M. R. et al. “The support of human genetic evidence for approved drug indications.” Nature Genetics 47, 856 to 860 (2015).

[11] Minikel, E. V. et al. “Refining the impact of genetic evidence on clinical success.” Nature 629, 624 to 629 (2024).

[12] Ioannidis, J. P. A. “Why Most Published Research Findings Are False.” PLoS Medicine 2, e124 (2005). doi:10.1371/journal.pmed.0020124. Argues from prior odds, statistical power and bias that a large share of published claims are false; the basis for treating the published literature as a noisy source of “known” causal biology.

[13] Okada, Y., Wu, D., Trynka, G., Raj, T. et al. (Plenge, R. M., senior author). “Genetics of rheumatoid arthritis contributes to biology and drug discovery.” Nature 506, 376 to 381 (2014). The empirical companion to the piece above: a genome-wide meta-analysis of over 100,000 subjects that recovered 18 of 27 approved RA drug targets from genetics alone, an enrichment significantly higher than the

equivalent enrichment for drug targets generally (P = 0.004).

[14] Rajagopal, V. “FNIP1 loss and protection from metabolic diseases.” GWAS Stories / LinkedIn (August 2026). A plain-language walkthrough of the FNIP1 paper's genetic-architecture logic.

[15] Trubetskoy, V., Pardiñas, A. F., Qi, T., Panagiotaropoulou, G., Awasthi, S. et al. (Schizophrenia Working Group of the Psychiatric Genomics Consortium). “Mapping genomic loci implicates genes and synaptic biology in schizophrenia.” Nature 604, 502 to 508 (2022). The PGC3 genome-wide association study of schizophrenia, mapping 287 loci and prioritizing candidate causal genes, including CHRM4, which encodes the M4 muscarinic receptor.

[16] Weiner, D. J., Nadig, A. et al. “Polygenic architecture of rare coding variation across 394,783 exomes.” Nature 614, 492 to 499 (2023). Quantifies rare coding-variant burden heritability at ~1.3% of phenotypic variance on average, and ~2% for schizophrenia and bipolar disorder, and shows common and rare variants converge on the same cell types while rare-variant signal concentrates in a tractable set of constrained, large-effect genes.

[17] Tian, R., Ge, T. et al. “Whole-exome sequencing in UK Biobank reveals rare genetic architecture for depression.” Nature Communications 15, 1755 (2024). Shows that the burden of rare damaging coding variants in loss-of-function-intolerant genes is significantly associated with depression risk, with rare-variant components overlapping autism, schizophrenia, and developmental disorders.

[18] Plenge, R. M. “Optimizing value creation in biopharma is like ‘the flop’ in Texas hold ‘em.” plengegen.com (2025). Develops end-to-end R&D principles, including matching the therapeutic modality to the biological mechanism of disease.

PART 2 OF 2

Causal Human Biology Is All You Need. But How Much Do We Know?

Why the Experiment Still Matters More Than the Model

Previously, in Part 1 (skip ahead if you've just read it)

Any piece of causal human biology sits in one of three tiers: known (established in the published literature), hidden (latent in data we've already generated but not yet extracted), or unknown (not yet revealed by any data, because the experiments haven't been run).

Which tier applies depends on the question. For drug discovery, there are three: the disease-level question (is this gene causal, and in which direction?), the mechanistic-driver question (through what pathway, in which tissue?), and the therapeutic-intervention question (can it be safely modulated for clinical benefit?).

Two lessons followed. Lesson #1: the answer depends on the application. Judged by clinical trial outcomes, we know roughly 10% of the causal biology relevant to drug discovery. **Lesson #2**: it varies sharply by disease area, from lipid disorders, where much is known, to psychiatric disease, where most is not.

One gene, *FNIP1*, stood out in a million-person study of lipid metabolism, with an effect size well beyond the rest. Part 2 starts there.

FNIP1, revisited: one gene, three questions, three different tiers

FNIP1 was one of 59 genes identified in the Regeneron Genetics Center's exome-wide association study looking at fat-storage capacity and it is a useful test case because its three questions land in three different tiers. Here is how each one played out.

The causal disease-level question sits firmly in the "known" tier. Heterozygous loss-of-function carriers show roughly 60% lower odds of a composite cardiometabolic outcome (odds ratio 0.39), and the relationship is graded rather than a single data point: complete biallelic loss causes a distinct, severe recessive syndrome, while heterozygous carriers, the population a drug would need to mimic, show no signal for that syndrome at all [5]. That is a textbook allelic series, exactly what a "known" causal answer looks like.

The causal mechanistic-driver question started in "unknown," and direct experimentation is what moved it into "known." A genetic association alone tells you *FNIP1* is causal; it does not tell you how, or where. That answer did not exist in any dataset until the authors generated it. Three experiments produced it.

First, they proposed the pathway: FNIP1 partners with folliculin (FLCN) to restrain two transcription factors, TFEB and TFE3, that drive mitochondrial and lysosomal biogenesis, so loss of FNIP1 should release those programs [5].

Second, they tested it directly in the causally relevant human tissue: silencing FNIP1 in primary human hepatocytes (over 90% mRNA knockdown) induced lysosomal and lipid-catabolism genes, exactly the direction the genetics predicted [5].

Third, they confirmed it *in vivo*: liver-selective ablation of the pathway in mice induced the full mitochondrial and lysosomal program and produced the metabolic benefit, protecting against diet-induced weight gain, lowering liver triglycerides, and improving insulin sensitivity [5].

Independently, tissue-enrichment analysis pointed to liver and adipose tissue as the sites where the causal biology plays out [5]. **None of this was extractable from existing data. It had to be run.** That is what it looks like when the mechanistic-driver question moves from unknown to known: not a better analysis of what we already had, but new experiments generating data that did not previously exist.

The causal therapeutic-intervention question sits in "unknown," and not simply because no drug has been tested. It sits there for two compounding reasons: (1) critical mechanistic details are still unresolved, and (2) no drug has yet been built and tested in humans. Consider what a serious FNIP1 drug discovery program would still need to answer. What modality perturbs this biology – an siRNA silencing FNIP1 selectively in hepatocytes, a small molecule disrupting the FLCN–FNIP1 interaction, or something else? Can any of those modalities achieve the tissue selectivity the safety case seems to require? Which node should you even hit: FNIP1, FLCN, or the complex itself? And what magnitude of perturbation is both safe and effective, given that heterozygous partial loss is protective while complete loss causes a severe recessive syndrome? A drug must land inside that therapeutic window, mimicking the carriers who benefit without approaching the phenotype of those who do not. None of those answers exist yet.

And the clearest reason to stay humble sits one gene away. FLCN, FNIP1's obligate binding partner in the very same complex, shows the identical protective direction of effect on the discovery phenotype. By the logic of the disease-level and mechanistic-driver questions alone, it should look like an equally promising target. It is not. Heterozygous FLCN loss causes Birt-Hogg-Dubé syndrome in humans, with an 18-fold higher risk of pneumothorax and a 9-fold higher risk of kidney cancer [5]. Two genes, one physical complex, the same causal direction, and completely different therapeutic answers. Unknown-unknowns clearly remain for this therapeutic hypothesis.

For a version of this same three-tier logic that did convert all the way into medicine, look one disease over. In inflammatory bowel disease the IL23R association (the causal disease-level question) [19], p19- versus p35-specific knockouts in mouse models of colitis (the mechanistic-driver question), and the head-to-head SEQUENCE trial (the therapeutic-intervention question) line up across all three tiers, and unlike FNIP1 they have already produced a better medicine: risankizumab, which targets the IL-23-specific p19 subunit, outperformed ustekinumab, which targets the p40 subunit shared by IL-12 and IL-23, most clearly on endoscopic remission, while IL-12-specific therapies never succeeded [20].

> **Lesson #3, and it builds directly on the previous two: Even for a single gene, "how much causal human biology do we know?" can have a different answer for each question**. FNIP1's disease-level question was answered by genetics alone, its mechanism needed new experiments, and its therapeutic question is still open. FLCN shows the risk of assuming otherwise: same complex, same protective direction, but a very different safety profile. A known answer to the first question should never stand in for the other two.

What AI can, and cannot, do about this

This blog's title borrows from a landmark 2017 AI paper, "*Attention Is All You Need*." Its claim was that attention, a mechanism that lets a model weigh which words in a passage matter most to one another, was enough on its own, without the more complicated machinery everyone had assumed was necessary [21]. My title makes a parallel argument for drug discovery: that the missing ingredient is not more computing power or bigger models, but causal human biology itself. Whether AI can deliver meaningful insight depends on the tier of knowledge. Where causal structure has already been pinned down cleanly, as in much of lipid biology, a model can learn from it directly.

But the analogy has a twist. Attention was only the mechanism for large language models. It worked because an enormous body of human writing already existed to learn from, a byproduct of everything people had put online. In that sense, causal biology is less like attention than like that body of text. Biology has no such byproduct: its causal facts exist only where someone has run the experiment to measure them. A better model can learn only from what has been measured, which is why generating that data, not just modeling it, is the binding problem.

AI has three jobs in R&D, and which one depends entirely on the tier of causal human biology. Where a piece of biology is **known**, AI's job is to compress and connect a well-instrumented pipeline. Where it is **hidden**, AI's job is to extract it from data that already exists: better models, more compute, better integration, aimed at what is already there. Where it is **unknown**, AI's job is to facilitate new discovery: generating hypotheses, then helping design and interpret the physical experiments needed to produce the data that does not exist yet. No amount of scale turns the third job into the first.

Demis Hassabis, who built both AlphaGo and AlphaFold, drew this line explicitly in a 2026 interview, tracing the extraordinary recent progress of AI on mathematics to a single property: **verifiability**. Proofs can be checked step by step, which lets a system try an idea, test it, learn, and iterate relentlessly. Biology is different. He expects AI to remain a collaborator rather than an autonomous discoverer in drug discovery, biology, and chemistry because “they're very messy, very emergent and you can't verify everything.” [22]

Daphne Koller of insitro sharpens the point. Agentic systems thrive where feedback is fast, cheap, and accurate, which is why AI has raced ahead on code and molecular design. But the actual causal quantity that matters in drug development – therapeutic benefit in a human being – can only be verified by a clinical trial: slow, expensive, and ethically bound in a way no amount of compute dissolves [3]. A 28-day toxicology study or a 52-week primary outcome in a clinical trial cannot be shortened to less than 28-days or 52-weeks, even with the most powerful AI capabilities. Common laboratory proxies, like liver-organoid cytotoxicity and thermal-shift binding assays, correlate poorly with the actual clinical outcomes they are meant to predict [1].

Sri Kosuri of Octant adds a further wrinkle for the unknown tier: what is missing there is not only the data but the question itself, which tends to arrive only once you start building. If you gave a scientist in the 1860s unlimited intelligence, they still could not have invented the transistor: the need for it had not yet emerged, the right questions had not yet been asked, and the methods to build it did not exist. Frontier drug discovery behaves the same way, in his account: you often know the target but not yet what a useful drug has to do, you lack the right assay, or you mistake an assay readout for the biology you actually care about. He expects AI's largest contribution to come from shortening the loop between a question and an experiment rather than from better foundation models alone, since models “can only learn as fast as they can explore the world” [23].

A 2025 Nature study shows what verifiable causal data can do and where it reaches its limits. By combining human loss-of-function "natural knockouts" and CRISPR-based Perturb-seq, the authors built an actual causal graph from regulator genes to biological programs to blood traits [7]. It worked, for three partially co-regulated blood traits in one well-instrumented cell line, and the authors are candid about why it stopped there: the signal vanished for traits governed by cell lineages their model did not cover, and they judge a complete, genome-scale causal network

"extremely challenging, if not infeasible" with today's data. That is the state of the art for AI-assisted causal discovery in human biology: real, and narrow.

The Bitter Lesson tension

In his 2019 essay "The Bitter Lesson," computer scientist Rich Sutton argued that in 70 years of AI history, general methods that scale with data and computing power have beaten approaches built on human expertise – in chess, vision, and speech alike [24]. For biology, the 2026 Cell perspective by Dupire and colleagues argues the reverse: AI models should be given what biologists already know, such as maps of which genes and proteins interact, rather than wait for scale to supply it [2].

Their argument rests on a shortage of data. The largest biological datasets contain roughly a thousandfold less data than the text used to train language models, and every data point must come from a costly experiment. Sutton's lesson assumes data can grow to match the problem. In biology, it can't yet. Dupire et al add that the knowledge they want to build into models reflects physical law, and they cite AlphaFold as proof that it works.

I think both sides are partly right, and the answer depends on the tier. Built-in knowledge pays off where biology is known or hidden, such as how a drug binds its target or how two proteins interact. AlphaFold succeeded because decades of protein structures were already banked in the Protein Data Bank. But for how whole cells and tissues behave, the knowledge on offer comes from curated interaction maps (STRING, ENCODE, ARACNe) that are incomplete, biased toward well-studied genes, and limited to the conditions in which they were measured. Relying on them risks locking in today's errors and blinding the model to the new causal links we most want to find. You cannot hand a model knowledge that does not yet exist. The dispute will be settled by how fast we can generate causal data relative to biology's complexity.

One more wrinkle: different kinds of biology probably need different kinds of models, such as network models for cell signaling and models of change over time for how cells respond to perturbation. One practical response is a system that routes each question to the most suitable method, as in Phylo's Biomni [25].

> **Lesson #4: What AI can do in a given area of biology depends on how verifiable the answers are, not on how sophisticated the model is.** Move-37-grade insight is possible only where results can be checked quickly and cheaply, as in math, Go, protein folding, and materials governed by known physics. In Go, the rules generate unlimited, perfectly reliable feedback. **Human disease biology has no such shortcut**. Reliable answers must come from data we already hold or from experiments not yet run, so they are limited by what has actually been measured. Even in the best-resolved areas, like lipid biology, the gap reopens when you ask a more specific question. Closing it is less a matter of scaling models than of collecting the right causal data: matched functional genomics, in the right cell types, at the scale human genetics now operates.

Where this leaves us

My conclusion is that **most causal human biology relevant to drug discovery is still unknown: not hidden in data we already hold, but absent from any dataset that exists, and reachable only through new experiments.** The evidence runs through the whole essay. Even for FNIP1, the mechanism and the therapeutic question could not be answered by analyzing existing data. And where verification is slow and costly, as in the brain, the gap is widest.

But “most” is not a number, and I opened this piece by demanding one from everyone else. So even though I’ve argued that a lone percentage is a category error, I owe you one anyway, if only so its structure can be unpacked rather than trusted.

> **My estimate:** In tissue-accessible domains with clean allelic series, like lipid metabolism, more than 20% of the clinically actionable causal biology is known with confidence. Even with this degree of knowledge, innovative medicines against genetically validated targets may still fail (e.g., Lp(a)). In diffuse, brain-centered domains with hard-to-define phenotypes, like psychiatric disease, it is below 10%. These are heuristic syntheses, not measurements, and both are read through germline genetics, so the low figure is a floor that broader data could raise.

The breakdown behind the number matters more than the number. The known tier is growing fastest, as biobank-scale genetics settles the disease-level question for more genes. The hidden tier is smaller, limited by tissue access and by how much existing data has been mined. The unknown tier is the slowest to shrink and the largest, though how much of it is truly unknown rather than merely hidden is the first open question below.

For anyone deciding where to place a near-term bet, whether capital, a research program, or trust in an AI system's output, four questions follow from the lessons:

1. What application are you optimizing for? (Lesson #1)
2. Which domain are you in, and how does it compare to others? (Lesson #2)
3. Which specific question are you asking? A strong genetic association earns real optimism on the disease-level question, cautious interest in the mechanism, and only a placeholder on the therapeutic one until it is separately earned. (Lesson #3)
4. Can the answer be checked quickly and cheaply, is it hidden in data we already hold, or must it be generated from scratch? (Lesson #4)

Apply the same discipline to AI. Ask whether a claimed discovery is compressing the literature, extracting something hidden, or driving a new experiment. Today the honest answer is usually the first, which is cause for optimism: the known and hidden tiers, where AI helps most, are expanding as biobanks, functional genomics, and real-world records are mined.

Open questions

I want to conclude not with a roadmap, such as that generated by Dupire and colleagues [2], but with three important open questions that will continue to shape the field.

How much of the unknown is hidden versus genuinely unknown, and can that split be estimated in advance rather than only in hindsight? The whole argument of this essay rests on a distinction it cannot yet quantify. Of the causal biology we don't currently understand, how much is hidden, latent in data we've already collected and recoverable by a capable enough model, versus unknown, absent from every existing dataset and reachable only by generating new data? A concrete test would be retrospective: for mechanisms resolved over the past decade, ask whether the deciding step was mining data that already existed or running an experiment that had to be done, with the criterion being what actually resolved each one, since in hindsight the old data almost always contains a trace. Even Causaly, whose product embodies the extraction position, concedes that the decisive step is often "the new observation that would distinguish" competing explanations [26]. For accessible, allelic-series-rich domains, extraction largely suffices. For much of the unknown tier, FNIP1 marks the limiting case: no retrieval layer can surface an experiment that was never run.

What other sources of causal human biology could sharpen the map, and does integrating them change the picture? Germline genetics is only one instrument, and the estimates in this essay are drawn almost entirely with it. Four other sources carry causal information it cannot see. Somatic mutations drive most cancers, where inherited variation explains only about 28% of risk [27]. Longitudinal profiling of patients before and after treatment speaks directly to the therapeutic question that genetics can only gesture at. Autoantibodies shape autoimmune disease, though in rheumatoid arthritis they sit atop a germline architecture that is already well resolved. Environmental exposures are often the proximal cause, as with Epstein-Barr virus in multiple sclerosis. These are parallel causal axes, not noise around the genetic signal. Integrating them would raise the estimate most where germline genetics understates it, as in cancer. The open question is whether that integration can be done systematically, rather than disease by disease in hindsight.

Could an AI world-model predict the result of an experiment before it is run, and how would we know it was good enough? Language and chemistry are relatively closed, low-entropy domains with dense, cheap, self-consistent data, part of why models learned them so well. Human biology is stochastic, radically diverse across individuals, and riddled with redundancy and context-dependence, so the question is not only whether such a model can be built but how much of biology's variance is learnable from data that already exists versus revealed only by doing the experiment. And “good enough” is the hard part of that question: for a given decision, a model that is right often enough to change what you fund or test next may matter more than a complete mechanistic account. Setting that bar, per decision and per disease, may be a more useful research program than chasing a universal one.

The Work Ahead

The recent leaps in AI were fueled by an enormous body of human writing that already existed. Biology has no equivalent. The causal facts that drug discovery needs, in the right tissues and at the right scale, mostly have not been generated, and no model, however capable, can learn from data no one has collected. AI will compress what we know and extract what is hidden, and both are worth doing. None of this argues for less AI in drug discovery; it argues for aiming it at the right failure. Most AI investment in the field goes to molecular design, while the dominant cause of clinical failure is the biology the molecule was aimed at. But the largest share of the work is still the experiments: finding out what we do not yet know, one question at a time.

Causal human biology may be all we need, but first we have to measure it.

Resources for further reading

Each entry is labeled with the citation number it corresponds to in the text above, and is listed in the order it first appears.

[5] Hindy, G., Adam, M. et al. “FNIP1 variants are associated with favourable metabolism in 1 million humans.” Nature (2026).

[19] Fachal, L. et al. “Resolving inflammatory bowel disease risk variants to genes and cell types.” medRxiv, posted 23 May 2026, doi:10.64898/2026.05.13.26352926. Preprint (not peer-reviewed). A multi-ancestry GWAS meta-analysis of 125,992 individuals with IBD and more than 1.2 million controls (619 independent association signals at 420 regions, 374 novel) that nominated 664 candidate effector genes and found genetically supported drug–target pairs were about 1.9-fold more likely to reach phase ≥3 trials than unsupported pairs (Fisher's exact P = 0.03).

[20] Peyrin-Biroulet, L., Chapman, J. C., Colombel, J.-F. et al. “Risankizumab versus Ustekinumab for Moderate-to-Severe Crohn's Disease.” New England Journal of Medicine 391, 213–223 (2024). doi:10.1056/NEJMoa2314585. SEQUENCE head-to-head phase 3b trial; anti–IL-23 (p19) risankizumab superior to anti-p40 ustekinumab on endoscopic remission at week 48 (31.8% vs. 16.2%).

[21] Vaswani, A. et al. “Attention Is All You Need.” Advances in Neural Information Processing Systems 30 (2017). The paper this piece's title deliberately echoes: the argument that a radically simplified architecture was the actual unlock for language modeling, not more of everything else.

[22] “Move 37.” Wall Street Journal (August 2026). Interview with Demis Hassabis on verifiability as the deciding variable behind AI's success in math and games versus its caution around biology; source of the “centaur era” framing.

[3] Koller, D. “Drug Discovery Has No Magic Wands.” a16z (August 2026). A founder's-eye view of the same argument, organized around disease-to-mechanism/mechanism-to-drug/drug-to-patient, evolutionary conservation as a predictor of tractability, and the feedback-loop-speed critique of agentic AI in biology.

[1] Bender, A. et al. “Artificial intelligence in drug discovery, what it is, where we stand and the path forward.” Nature Reviews Drug Discovery (2026). The most comprehensive recent critique of AI's translational impact; the source of the proxy-data figures (liver-organoid DILI, thermal-shift binding) and the reductionism/emergence argument via Lazebnik and Anderson.

[23] Kosuri, S. “Mens et Manus.” Octant / Substack (srikosuri.substack.com/p/mens-et-manus). Argues from inside a drug-discovery company that building is how understanding is produced rather than what follows it, using the transistor and Bell Labs' failed solid-state amplifier as the worked example, and concludes that AI's largest contribution to frontier drug discovery will come from shortening the loop between question and experiment rather than from better foundation models alone.

[7] Ota, M. et al. “Causal modelling of gene effects from regulators to programs to traits.” Nature 650, 399 to 408 (2025). The current state of the art in combining human loss-of-function data with Perturb-seq to build genuine causal graphs: narrow, well-instrumented, and explicit about the limits of scaling further.

[24] Sutton, R. “The Bitter Lesson.” incompleteideas.net (13 March 2019). The influential essay distilling seventy years of AI history into a single claim, that general methods leveraging computation and data ultimately outperform approaches built on human-engineered domain knowledge; the counterpoint the Cell perspective sets out to answer for biology.

[2] Dupire, L., Khan, A. A., Karaletsos, T., Kelley, S., Lundberg, E., Ma, J., Paull, E., Quake, S. R., Rabadan, R., Rowan, C., Sims, P., Tavazoie, S., Tsang, J. S., Zhang, M. & Califano, A. “Fifteen challenges for generative AI applications to cell biology.” Cell 189 (September 17, 2026). doi:10.1016/j.cell.2026.07.004. A Leading Edge Perspective proposing fifteen disease-agnostic grand challenges for generative AI in cell biology, most of which still lack effective predictive methodologies, and arguing that insufficient measurement data and multicellular complexity, not compute, are the binding constraints.

[25] Huang, K., Zhang, S., Wang, H., Qu, Y. et al. “Autonomous biomedical research with an artificial intelligence agent.” Science 393 (2026). Describes Biomni, a general-purpose biomedical AI agent from Phylo that composes specialized tools, databases, and code execution into research workflows; the tool-calling-harness example cited in the text.

[26] Kiachopoulos, Y. “Building the AI Co-Scientist for the Hardest Part of Science.” Causaly (July 21, 2026).

[27] Lichtenstein P, Holm NV, Verkasalo PK, et al. Environmental and heritable factors in the causation of cancer — analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med. 2000;343(2):78–85.