

Data science in drug discovery safety: Challenges and opportunities

Nicholas J Coltman¹, Ruth A Roberts^{1,2}  and James E Sidaway¹

¹Apconix, Alderley Edge, Cheshire SK10 4TG, UK; ²School of Biosciences, University of Birmingham, Birmingham B15 2TT, UK
Corresponding author: Ruth A Roberts. Email: ruth.roberts@apconix.com

Impact Statement

Target safety assessments (TSAs) are essential to provide drug projects with the best chance of success. Data science, if used correctly, offers a powerful tool for accurate and informative TSAs.

Abstract

Early de-risking of drug targets and chemistry is essential to provide drug projects with the best chance of success. Target safety assessments (TSAs) use target biology, gene and protein expression data, genetic information from humans and animals, and competitor compound intelligence to understand the potential safety risks associated with modulating a drug target. However, there is a vast amount of information, updated daily that must be considered for each TSA. We have developed a data science-based approach that allows acquisition of relevant

evidence for an optimal TSA. This is built on expert-led conventional and artificial intelligence-based mining of literature and other bioinformatics databases. Potential safety risks are identified according to an evidence framework, adjusted to the degree of target novelty. Expert knowledge is necessary to interpret the evidence and to take account of the nuances of drug safety, the modality, and the intended patient population for each TSA within each project. Overall, TSAs take full advantage of the most recent developments in data science and can be used within drug projects to identify and mitigate risks, helping with informed decision-making and resource management. These approaches should be used in the earliest stages of a drug project to guide decisions such as target selection, discovery chemistry options, *in vitro* assay choice, and end points for investigative *in vivo* studies.

Keywords: Drug safety, target safety, bioinformatics

Experimental Biology and Medicine 2023; 248: 1993–2000. DOI: 10.1177/15353702231215890

Introduction

Drug discovery and development is a complex process requiring considerable investment of time and money. Programs usually begin with the selection of a target based on target linkage to disease and assessment of likely “drugability.”¹ Following this, a decision may be taken on whether to proceed with a traditional small molecule approach, a biologic (antibody or fragment) or to pursue one of the many new modalities under development such as antisense oligonucleotides or targeted protein degraders.

Much emphasis is placed within projects on efficacy and target linkage to disease. However, since most projects failures are due to safety, and some 25–50% of these are due to the drug target itself,² it is imperative to make safety part of drug design. Each modality has its own advantages and risks from a safety point of view. For example, small molecules can be associated with off-target chemical toxicity such as hERG liability or inhibition of hepatic drug transporter proteins. Similarly, antibodies and oligonucleotides might be associated with immunogenicity and/or nephrotoxicity

(see Figure 1). As well as drug modality-related toxicity, it is also imperative to characterize the potential for unwanted side effects attributable to the target itself.³ Specifically, each drug target may have on-target but off-tissue toxicities that can be attributed to exaggerated pharmacology and/or a different action in a different tissue; this may arise when the intended target is expressed in both diseased and healthy tissue. Previous publications have described the value of target safety assessments (TSAs) to drug projects, with particular emphasis on saving time and resources via early mitigation of risk.^{3,4} Here, we will consider the generation and use of TSAs for decision-making in drug projects.

Case study 1: TSAs, when and why?

TSAs are preferably carried out early in drug discovery and development but may be carried out at any stage (Figure 2).³ During target selection, TSAs are useful for identifying showstoppers and for comparing targets as part of decisions to accept a new target into the portfolio.⁴ Modality can also be considered at this stage; for example, how would the

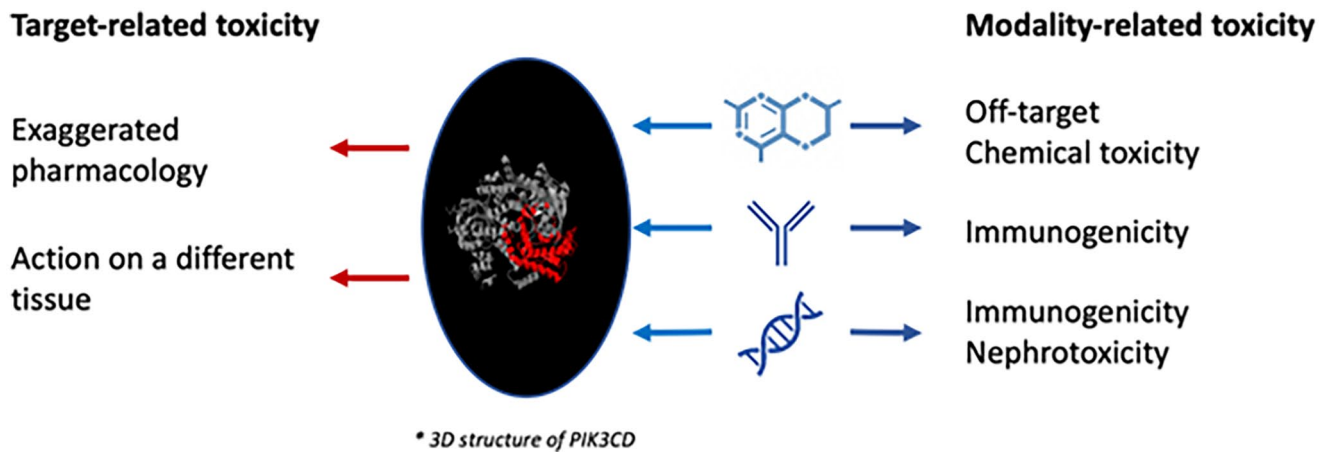


Figure 1. Making safety part of drug design. Safety is the main reason for failure in drug discovery and development. The development of “safer” drugs requires an in-depth knowledge of potential adverse effects related to the biology of the drug target as well as toxicity risks that are carried by the drug modality themselves.

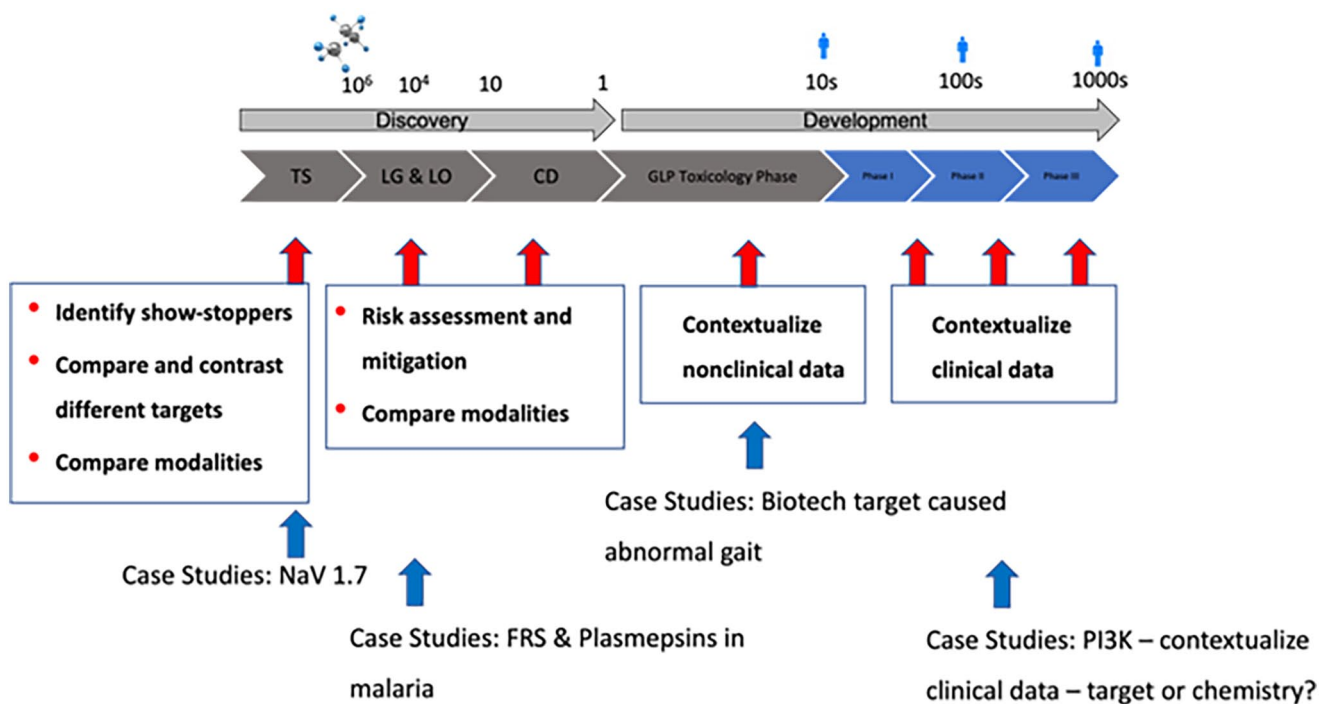


Figure 2. Target safety assessments (TSAs) during drug discovery and development. TSAs may be carried out throughout the process from target selection (TS) to clinical development.

Source: Figure adapted from Walker *et al.*¹

LG: lead generation; LO: lead optimization; CD: candidate drug; GLP: good laboratory practice.

potential safety profile of an antibody compare with that of a small molecule against the same target? Once targets are accepted into the portfolio, a more comprehensive TSA during lead optimization and candidate drug selection is very valuable to provide a risk assessment and a risk mitigation plan. TSAs can also be used to place non-clinical and clinical data into context. With the increasing numbers of tractable drug targets and modality types, getting to know your target biology and gaining a deeper insight to the possible mechanisms at play with modulation of your drug target, the TSA is a principal tool in the armament of toxicologists in drug development.⁵

Several case studies are highlighted in Figure 2. NaV1.7 is a voltage-gated sodium channel preferentially expressed in neurons and modulation of NaV1.7 through inhibition is a potential mechanism to temper pain sensitivity.⁶ Phenotypic outcomes and predictive toxicity from human and animal genetic studies of NaV1.7 are likely to be limited to anosmia (loss of sense of smell) due to expression of the target in olfactory sensory neurons – an acceptable side effect for this therapeutic strategy. However, there are many closely related sodium channels with much more serious potential side effects such as Long QT syndrome or seizure disorders, so it was pivotal for this drug project to achieve a high degree

of specificity for the intended target over other potential off-target proteins.⁷

Plasmeprins are important antimalarial drug targets due to their specificity for the malaria parasite and their vital role as mediators of disease progression.⁸ However, there is a risk of toxicity if there is insufficient selectivity over related mammalian aspartic proteases. For this project, it was critical to assess potential inhibitors of malarial plasmeprins against the main human aspartic proteases and especially CatD/E.⁸ Assuming adequate target homology and mechanistic translation between human and rodent species, an investigative rodent study conducted early was advised to ensure an early risk assessment of the hazards identified for both the NaV1.7 and Plasmeprin projects.

For a biotechnology project, an abnormal gait was noticed in rats during Good Laboratory Practice (GLP) toxicology after only seven days of dosing. Several reasons were proposed for this such as central nervous system (CNS) or muscle effects. However, an evaluation of the target suggested it was expressed in bone and was involved in bone growth, perhaps explaining what was seen in the animals. Importantly, the toxicological effect was predicted only to occur in adolescents where the bone is still growing but not in adults. Unfortunately, for this project, the study animals were not yet fully mature, explaining the findings seen. A TSA could have predicted this risk and avoided generating data not relevant to the clinical population. In one final example, findings were noted in the clinic for a small molecule kinase inhibitor, but their origin was unclear. An evaluation of the chemistry did not provide any clues as to the origin of the clinical finding, but a TSA indicated that the findings were likely to be due to the target rather than the chemistry. This important finding meant that the project could not avoid unwanted toxicity via different chemistry but rather needed to re-evaluate the tractability of the target in the current indication.

Generating a TSA

There are three key steps to generating a TSA: identifying evidence sources, determining which evidence is relevant, and then evaluating all the evidence to conduct the risk assessment. Evidence sources rely primarily on the literature, and full-text analyses are absolutely critical as key safety information may not be absent from the abstract. Moreover, even mining full-text documents can often be hampered by lack of optical character recognition (many toxicology studies on approved drugs may have been published several decades ago and may only exist as photocopied evidence). Similarly, conference reports and preprints can be useful to find the most recent information on a target. Bioinformatics databases are also used and information about assets that engage the target can be accessed from an ever-growing number of resources such as IUPHAR⁹ or Open Targets,¹⁰ or competitor intelligence sources.

Once evidence is assembled, the next step is to identify the evidence that is most relevant for inferring safety risks. In this, mouse genetic data such as knockouts and overexpression are very useful for novel targets, depending on whether

the intention is drug-target inhibition or activation.^{11,12} Similarly, human genetic conditions where the intended target is deleted or modified can give insight into the likely outcomes of target modulation; several databases such as OMIM¹³ and recent advances in GWAS have been used to support target safety by leveraging human genetic data.¹⁴ Animal and clinical data from previous drugs that interact with the target might exist for more mature targets especially if the drug has reached the market and the drug approval information is in the public domain. Other than this, there may be publications of toxicology data especially for drug programs that have failed. These data on previous drugs that have made it to preclinical or clinical studies are very useful, providing there is some insight into what is likely to be target related rather than due to the modality. In this, class effects or outcomes that are corroborated by other evidence types are very useful.

Once relevant evidence is identified, a risk assessment is conducted where potential target-related toxicities are identified across each organ system via a synthesis of all the different evidence sources.¹⁵ Critically, data can be weighted by an expert toxicologist to give an insight into the likelihood of a particular outcome based on the weight of evidence and an assessment of impact for the project if the outcome was to occur. For example, one or two literature sources reporting *in vitro* data are likely to offer a lower weight of evidence compared with reports from human clinical trials. Nonetheless, these *in vitro* data might provide valuable mechanistic insight alongside genetic-phenotypic studies. Similarly, as described earlier, anosmia as a result of NaV1.7 modulation might pose a low risk for the progression of a chronic pain project whereas risk of seizure would pose a high risk.

Case study: the use of evidence sources – advantages and potential pitfalls

To illustrate how evidence sources can be used in a TSA, we aligned gene and protein expression data from the Genotype Expression Project¹⁶ and the Human Protein Atlas¹⁷ for phosphoinositide 3-kinase delta (PI3K δ), a well-founded and Food and Drug Administration (FDA)-approved target in oncology (Figure 3). We concatenated the full GTEx and human protein atlas (HPA) whole tissue RNA-Seq data equating in some instances to more than $n = 100$ samples per tissue, allowing the full distribution of data to be studied. Together with single-cell RNA-Seq (scRNA-Seq) and immunohistochemistry data from the HPA, we aligned these with the common organ systems routinely described by pathologists and toxicologists.¹⁸ Alignment of the three data sets as a tryptic plot is very useful for overcoming some of the limitations of each individual platform. For example, mRNA expression is unlikely to correlate with protein, and for immunohistochemistry antibody reliability is a constant concern. Whole tissue RNA-Seq or proteomics can mask high expression in individual cell types, but this can be overcome by scRNA-Seq data if available. In this example, there is high expression of PI3K δ in the immune system and resident immune cells such as the microglia in the nervous

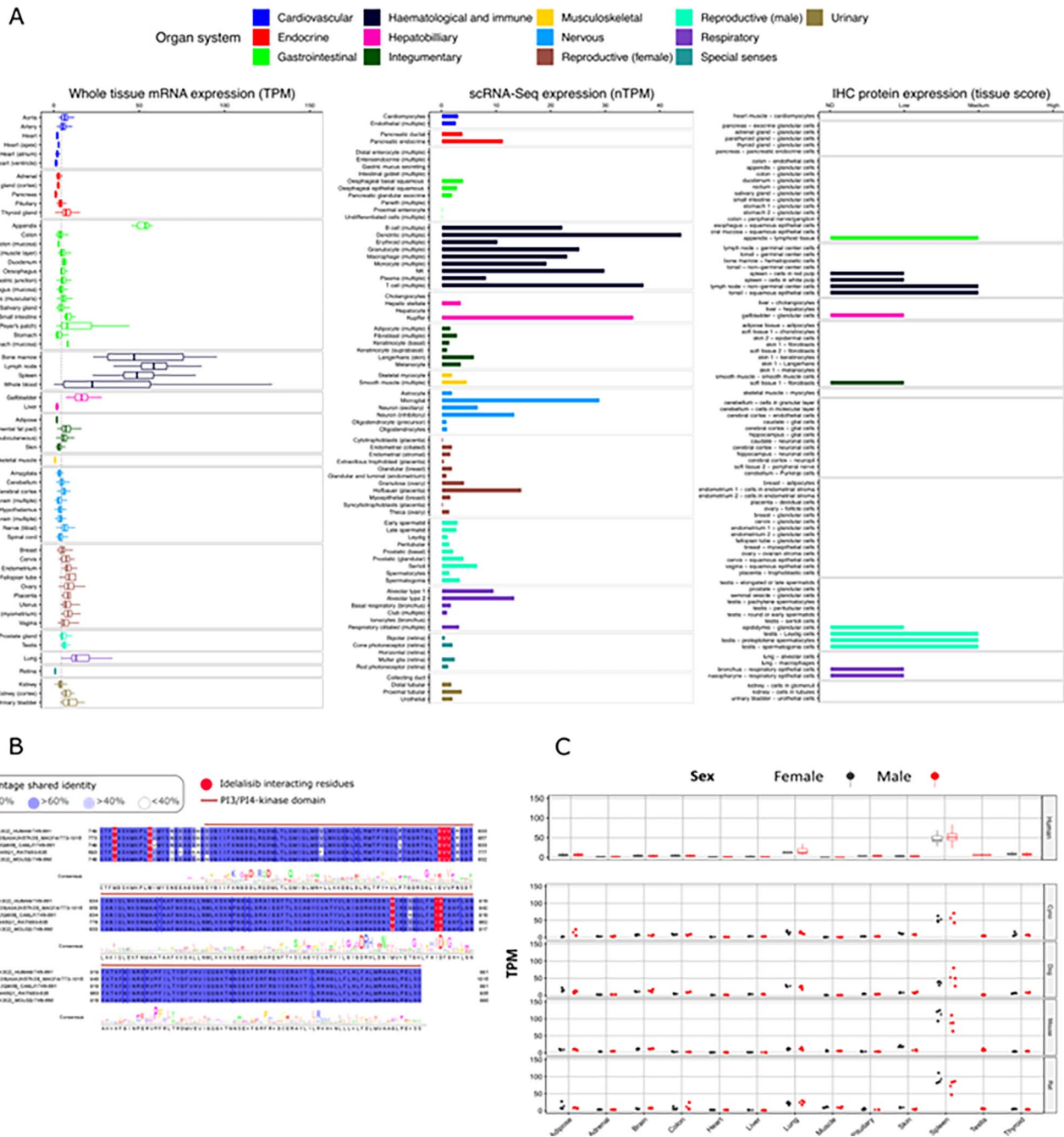


Figure 3. The Use of Bioinformatics to study target distribution and translatability to animal species used in toxicology studies. Plots show gene and protein expression data for PI3K δ . (A) The expression profile of PI3K δ was studied at both the mRNA (whole tissue and single-cell RNA-Seq) and protein level (immunohistochemistry data), aligned with key target organ systems. RNA-Seq data are curated from GTEx and HPA. Immunohistochemistry data are from all tissues assessed by HPA and scored according to the HPA methodology. The PI3K δ antibody used by HPA was considered to be of good reliability. (B) Alignment of the PI3K δ PI3/P14 kinase domain was performed between human, cynomolgus monkey (MACFA), dog (CANLF), rat, and mouse using Clustal OWS. Residues are colored according to the percentage of residues in each column that agree with the consensus sequence (percentage shared identity). Only the residues that agree with the consensus residue for each column are colored. Residues bound by FDA-approved PI3K δ inhibitors idelalisib, duvelisib, and umbralisib are highlighted in red.²³ The motif to depict the Hidden Markov Model was derived from seed alignments curated by Pfam for the PF00454 family (41 seed sequences). (C) The mRNA expression profile of PI3K δ was compared by RNA-Seq between human (GTEx) and the main toxicology species matched normal tissues. Expression plotted as transcripts per million (TPM), across sample types and organized by sex. Preclinical species RNA-Seq expression data were obtained from PRJNA516470.²⁴

system and the Hoffbauer cells in the placenta. However, this was not reported in the immunohistochemical data set. Overall, this type of in-depth analysis allows potential target organs and cell types to be identified for consideration of target related toxicities. This is especially useful if the modality is an antibody drug conjugate for consideration of tissue

cross reactivity. In addition, for many targets, it can suggest tissue sites for potential target-related toxicity for deeper literature mining, or it might indicate a wide range of tissue safety risks for a ubiquitous target. Translatability of target expression and function between preclinical species is vital for successful drug development.

Thus, one key aim of the TSA is to identify instances of high or low genetic conservation and mechanistic translatability. For example, ideally, there is strong homology of a drug target between humans and the likely species to be used in toxicology and efficacy studies, but this is not always the case. Purinergic receptors P2X2 and P2X3 are two drug targets implicated in chronic pain, for which a mechanistic basis has mostly been derived from use of rodent models.¹⁹ However, there are significant species differences between rodents, non-human primates, and humans in pain mechanisms. Therefore, the translation of the efficacy of target antagonism across species is challenging and requires a careful evaluation of species differences in isoforms and expression. Undertaking a TSA could have likely identified these translation challenges earlier because not only do the mRNA expression profiles of P2X2 and P2X3 differ between rodents and primates but the binding domains and amino acid residues also differ between rats and humans, thus impacting the potential binding of a P2X3 antagonist and decreasing potency. The TSA therefore would highlight that the rodent is not an appropriate species for investigative work in this case. Similarly, a single amino acid substitution in an ATP binding pocket of tyrosine kinase 2 (TYK2) – frequently targeted for autoimmune conditions like arthritis and psoriasis – was enough to see a significant shift in potency of a novel TYK2 inhibitor, between humans and preclinical species.²⁰ Conducting sequence alignments of TYK2 between human, non-human primates, rodent, and canine species within a TSA would have likely identified these key differences.

Case study: use of bioinformatics in translational assessment

One of the key challenges in a drug project is ensuring efficacy and safety data are relevant for humans. For example, a lead molecule might bind to the mouse target protein and show good efficacy in mouse models but subsequently this implied efficacy might not translate to humans. Similarly, there could be examples where toxicology data generated in animals are not relevant for human safety due to differences in drug-binding domains and/or to target expression. It would surely be far better to understand these cross-species differences at the start of a project rather than trying to back-rationalize confusing data. Figure 3(B) shows a high degree of similarity in the drug-binding domain of PI3K δ between mouse (commonly used for efficacy), common preclinical toxicology species and humans, giving confidence in the translation of data across species. Similarly, the mRNA expression levels of the target across humans and preclinical species (Figure 3(C)) shows high similarity across 13 matched tissues. Together with genetic conservation of the key drug-binding domain, these data provide some reassurance that the target is likely to respond in a comparable manner with respect to pharmacology and tissue specificity and that mechanistic biology is likely to be retained between preclinical species and humans. However, as ever, it would always be important to assess the literature for reports for other relevant information on potential species differences.

Case study – the limitations of using precurated data for TSAs

As the wealth of scientific literature and data related to drug discovery has increased, so too have the number of ways that these rich sources can be accessed and mined. Resources such as HPA, Mouse Genome Index (MGI, Blake *et al.*¹²), Open Targets,⁹ and other initiatives, are all valuable, information-rich knowledgebases that could be leveraged to support TSAs alongside data mined from the literature. HPA offers a useful snapshot, as well as in-depth analyses of human protein organ, tissue, and cellular distribution that could be used to understand target expression in drug discovery projects. However, much of these data are curated from numerous other databases and literature sources which means that not all findings reflected in the original data source are necessarily translated into the HPA database and are potentially prone to errors during the data curation and interpretation process. In some cases, data might be reanalyzed and aligned to an in-house bioinformatics pipeline or simply just made available for use which may give rise to discrepancies between interacting data sets.

In one case study, HPA (version 22) reported our target of interest as having high mRNA expression in resident hepatic macrophages (Kupffer cells) by scRNA-Seq which was at odds with our knowledge of the target and the literature. Moreover, it was not clear where this potential error had stemmed from. To address this problem, we mined the literature and accessed original data sources to derive a logical conclusion rather than that assumed *a priori* from HPA. From the literature, we identified two key marker genes that could separate proinflammatory hepatic macrophages from tolerogenic macrophages (VCAN and MARCO, respectively^{21,22}) and compared their expression patterns (Figure 4). Overall, the absolute mRNA expression (normalized Transcript per Million, nTPM) of the target of interest across various macrophage populations in multiple tissues was significantly lower than that of VCAN and MARCO macrophage marker genes. Our target was found to have enhanced expression at the HPA single-cell level in multiple macrophages, dendritic cell, Langerhans cell, and monocytes populations, but unexpectedly showed the highest expression in hepatic Kupffer cells. MARCO, a marker of tolerogenic macrophages, had high expression in general macrophage populations but to a lesser extent in Kupffer cells. VCAN, a marker of inflammatory macrophages, showed the highest expression in Kupffer cells, but a lower expression in macrophages or monocytes, in-line with our knowledge of these cell types and marker genes (see Supplemental data).

On further analysis of the metadata and literature, it appears that one analysis²¹ does not curate Kupffer cells directly, but instead labels these cells as “non-inflammatory macrophages based on their similarity to mouse KC.” By comparison to marker genes differentially expressed in inflammatory or tolerogenic hepatic macrophages (Figure 4), the clustering of data indicates that the target of interest had a cellular phenotype more closely resembling that of cluster 4 (inflammatory macrophage) and not a Kupffer cell, *per se*. Overlaying our target expression across the full

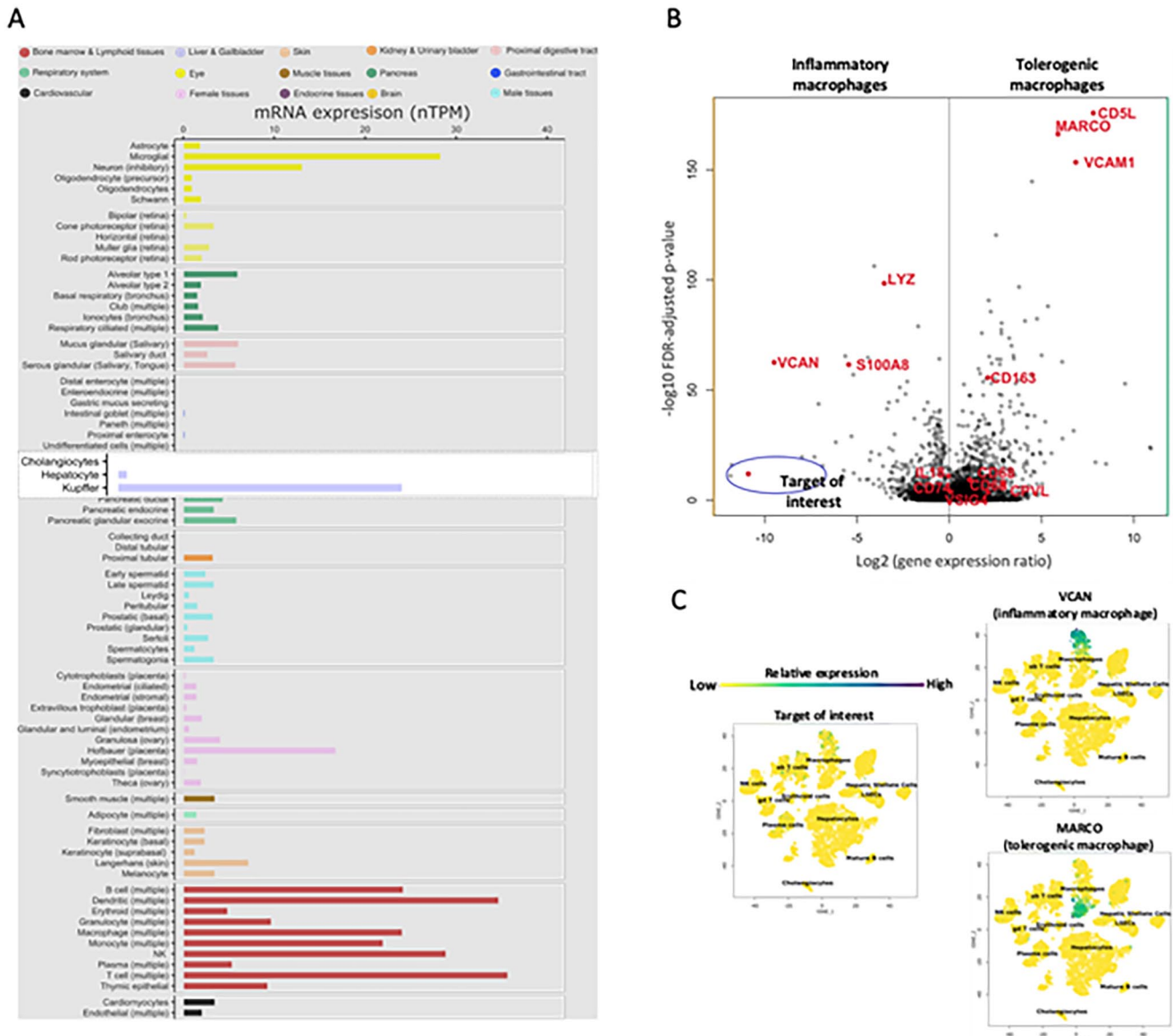


Figure 4. Single-cell RNA-Seq expression analysis of the human liver reveals that the target of interest expression is negligible in hepatic macrophages. (A) Expression of the target of interest in hepatic macrophages was compared using the data originally derived from MacParland *et al.*²¹ before curation by HPA consortium. (B) Differentially expressed genes (DEGs) with pairwise gene expression ratio were compared between inflammatory hepatic macrophage and tolerogenic macrophage clusters and (C) dimension reduction of cell distribution of the target of interest compared to markers of tolerogenic (MARCO) and inflammatory macrophages (VCAN) performed by tSNE. Distribution of hepatic cells is depicted as clusters pertaining to the principal hepatic cell types; color bar overlay represents a scale of target expression (low to high).

hepatic cell-type map indicates that expression of our target is absent in all cell-type clusters analyzed whereas MARCO and VCAN, markers of two hepatic macrophage subsets, showed isolated expression in two macrophage clusters. In the original paper,²¹ prior to curation by HPA, the preparation and methodology to derive hepatic cell homogenates were noted as a potential caveat for interpretation of results from scRNA-Seq analysis. As such, it would seem reasonable to assume that the viability and heterogeneity of hepatic cells sorted for analysis by the original authors may not be a true biological reflection of the native liver biology. The authors note that although the five caudate liver lobes collected for sampling were deemed “clinically acceptable,” they were

nevertheless obtained from neurologically dead patients and did exhibit mild inflammation.

For further confirmation, we analyzed data from second single-cell liver data sets in cell subsets taken from six healthy liver donors.²³ This revealed two main subsets of Kupffer cells (cluster 6: LIRB5⁺CD5L⁺MARCO⁺HMOX1^{high} and cluster 2: CD1C⁺FCER1A⁺). On inspection of the data, we found a general absence of the target of interest in the 39 cell clusters analyzed in the sc-RNA-Seq data set, again showing that our target did not cluster with any hepatic cell type, including subsets of Kupffer cells.

There is therefore the possibility that the ratio of pro-inflammatory versus tolerogenic hepatic macrophage

populations was imbalanced in the original analyses. The single-cell analysis conducted represents populations of hepatic cell types but not necessarily their actual frequency within the original healthy tissue. In conclusion, potential errors in the curation of the HPA scRNA-Seq data set could have led to the wrong conclusions concerning the expression of our target, potentially impacting data interpretation on the risk of off-tissue but on-target engagement in the liver.

Conclusions

Overall, it is imperative to make safety part of drug design by considering potential toxicities related to the chemistry/modality as well as the unintended consequences of modulating the target. TSAs are evolving constantly to take full advantage of the most recent developments in data science, enabling the identification and mitigation of risks within drug projects. Alignment and expression data are vital in understanding the translation from preclinical species to humans, both with respect to efficacy and safety. Together with an assessment of modality, these assessments are used to drive informed decision-making and resource management. These approaches should be used in the earliest stages of a drug project to guide decisions such as target selection, discovery chemistry options, *in vitro* assay choice, and end points for investigative *in vivo* studies.

AUTHORS' CONTRIBUTIONS

NJC, RAR, and JES contributed equally to this manuscript.


DECLARATION OF CONFLICTING INTERESTS

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: RR is co-founder and co-director of Apconix, an integrated toxicology and ion channel research company that provides expert advice on non-clinical aspects of drug discovery and drug development to academia, industry, government, and not-for-profit organizations. JES and NJC are the employees of Apconix.

FUNDING

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This minireview highlights previously published data and concepts. There was no specific funding required for this work so there is none to report.

ORCID ID

Ruth A. Roberts  <https://orcid.org/0000-0002-7763-7558>

SUPPLEMENTAL MATERIAL

Supplemental material for this article is available online.

REFERENCES

- Walker AL, Imam SZ, Roberts RA. Drug discovery and development: biomarkers of neurotoxicity and neurodegeneration. *Exp Biol Med* 2018;**243**:1037–45
- Cook D, Brown D, Alexander R, March R, Morgan P, Satterthwaite G, Pangalos MN. Lessons learned from the fate of AstraZeneca's drug pipeline: a five-dimensional framework. *Nat Rev Drug Discov* 2014;**13**:419–31
- Roberts RA. Understanding drug targets: no such thing as bad news. *Drug Discov Today* 2018;**23**:1925–8
- Brennan RJ. Target safety assessment: strategies and resources. *Methods Mol Biol* 2017;**1641**:213–28
- Blomme EAG, Will Y. Toxicology strategies for drug discovery: present and future. *Chem Res Toxicol* 2016;**29**:473–504
- Emery EC, Luiz AP, Wood JN. Nav1.7 and other voltage-gated sodium channels as drug targets for pain relief. *Expert Opin Ther Targets* 2016;**20**:975–83
- Sadler C, Eades L, Knight R, Morton M, Roberts RA. Target Safety Assessments and their role in de-risking drug discovery: Nav1.7 as an example, 2017, <https://www.apconix.com/wp-content/uploads/2021/05/Target-safety-assessments-and-their-role-in-de-risking-drug-discovery-Nav1.7-as-an-example.pdf>
- Barber J, Sikakana P, Sadler C, Baud D, Valentin JP, Roberts R. A target safety assessment of the potential toxicological risks of targeting plasmeprin IX/X for the treatment of malaria. *Toxicol Res* 2021;**10**:203–13
- Harding SD, Armstrong JF, Faccenda E, Southan C, Alexander SPH, Davenport AP, Pawson AJ, Spedding M, Davies JA, NC-IUPHAR. The IUPHAR/BPS guide to PHARMACOLOGY in 2022: curating pharmacology for COVID-19, malaria and antibacterials. *Nucleic Acid Res* 2022;**50**:D1282–94
- Ochoa D, Hercules A, Carmona M, Suveges D, Baker J, Malangone C, Lopez I, Miranda A, Cruz-Castillo C, Fumis L, Bernal-Llinares M, Tsukanov K, Cornu H, Tsigirgos K, Razuvayevskaya O, Buniello A, Schwartzenuber J, Karim M, Ariano B, Martinez Osorio RE, Ferrer J, Ge X, Machlitt-Northen S, Gonzalez-Urriarte A, Saha S, Tirunagari S, Mehta C, Roldán-Romero JM, Horswell S, Young S, Ghousiani M, Hulcoop DG, Dunham I, McDonagh EM. The next-generation Open Targets Platform: reimaged, redesigned, rebuilt. *Nucleic Acid Res* 2023;**51**:D1353–9
- Bhoumik P, Del Rio-Espinola A, Hahne F, Moggs J, Grenet O. Translational safety genetics. *Toxicol Pathol* 2017;**45**:119–26
- Blake JA, Baldarelli R, Kadin JA, Richardson JE, Smith CL, Bult CJ, Mouse Genome Database Group. Mouse Genome Database (MGD): knowledgebase for mouse-human comparative biology. *Nucleic Acid Res* 2021;**49**:D981–7
- McKusick-Nathans Department of Genetic Medicine. OMIM®: an online catalog of human genes and genetic disorders. 2023, <https://omim.org/>
- Carss KJ, Deaton AM, Del Rio-Espinola A, Diogo D, Fielden M, Kulkarni DA, Moggs J, Newham P, Nelson MR, Sistare FD, Ward LD, Yuan J. Using human genetics to improve safety assessment of therapeutics. *Nat Rev Drug Discov* 2022;**22**:145–62
- Sadler C, Roberts RA. What are the benefits of a target safety assessment? 2023, <https://www.ddw-online.com/how-a-target-safety-assessment-supports-the-drug-development-process-2-24888-202307/>
- GTE Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* 2013;**45**:580–5
- Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson Å, Kampf C, Sjöstedt E, Asplund A, Olsson I, Edlund K, Lundberg E, Navani S, Szgyarto CAK, Odeberg J, Djureinovic D, Takanen JO, Hober S, Alm T, Edqvist PH, Berling H, Tegel H, Mulder J, Rockberg J, Nilsson P, Schwenk JM, Hamsten M, von Feilitzen K, Forsberg M, Persson L, Johansson F, Zwahlen M, von Heijne G, Nielsen J, Pontén F. Proteomics. Tissue-based map of the human proteome. *Science* 2015;**347**:1260419
- Sidaway J, Sikakana P, Haynes B, Ge H, Roberts R. SOC-III-05 What are the common predicted toxicities from target safety assessments? *Toxicol Lett* 2022;**368**:S44
- Serrano A, Mo G, Grant R, Paré M, O'Donnell D, Yu XH, Tomaszewski MJ, Perkins MN, Séguéla P, Cao CQ. Differential expression and pharmacology of native P2X receptors in rat and primate sensory neurons. *J Neurosci* 2012;**32**:11890–6
- Gerstenberger BS, Banker ME, Clark JD, Dowty ME, Fensome A, Gifford R, Grifford MC, Hegen M, Hollingshead BD, Knafels JD, Lin TH, Smith JF, Vajdos FF. Demonstration of *in vitro* to *in vivo* translation of

- a TYK2 inhibitor that shows cross species potency differences. *Sci Rep* 2020;**10**:8974
21. MacParland SA, Liu JC, Ma XZ, Innes BT, Bartczak AM, Gage BK, Manuel J, Khuu N, Echeverri J, Linares I, Gupta R, Cheng ML, Liu LY, Camat D, Chung SW, Seliga RK, Shao Z, Lee E, Ogawa S, Ogawa M, Wilson MD, Fish JE, Selzner M, Ghanekar A, Grant D, Greig P, Sapi-sochin G, Selzner N, Winegarden N, Adeyi O, Keller G, Bader GD, McGilvray ID. Single cell RNA sequencing of human liver reveals distinct intrahepatic macrophage populations. *Nat Commun* 2018;**9**:4383
 22. Aizarani N, Saviano A, Mailly L, Durand S, Herman JS, Pessaux P, Baumert TF, Grün D. A human liver cell atlas reveals heterogeneity and epithelial progenitors. *Nature* 2019;**572**:199–204
 23. Vanhaesebroeck B, Perry MWD, Brown JR, André F, Okkenhaug K. PI3K inhibitors are finally coming of age. *Nat Rev Drug Discov* 2021;**20**:741–69
 24. Naqvi S, Godfrey AK, Hughes JF, Goodheart ML, Mitchell RN, Page DC. Conservation, acquisition, and functional impact of sex-biased gene expression in mammals. *Science* 2019;**365**:eaaw7317