

Reviewed Preprint

v1 • February 19, 2026

Not revised

For correspondence:

beloeillrt@gmail.com

Competing interests: No

competing interests declared

Reviewing editor: Gyaviira

Nkurunungi, Medical Research Council/Uganda Virus Research Institute and London School of Hygiene and Tropical Medicine Uganda Research Unit, Uganda

© 2026, Becker et al. This article is distributed under the terms of the

[Creative Commons Attribution](#)

[License](#), which permits unrestricted use and redistribution provided that the original author and source are credited.

A translational transcriptomic signature of vaccine reactogenicity for the evaluation of novel formulations

Jérémie Becker, Maroussia Roelens, Kendra Reynaud, Laurent Beloeil 

BIOASTER, Lyon, France

eLife Assessment

This study proposes a cross-species transcriptomic framework to predict vaccine reactogenicity, with implications for preclinical vaccine safety assessment. The findings show that mouse muscle transcriptomic signatures capture conserved inflammatory programs and can identify highly reactogenic formulations, with supportive but limited evidence for finer discrimination among licensed human vaccines. Overall, the work establishes a **valuable** foundation for translational biomarkers of reactogenicity, although the strength of evidence for broad cross-species predictive performance remains **incomplete** and would benefit from further validation.

<https://doi.org/10.7554/eLife.109928.1.sa2>

Abstract

Accurately predicting vaccine reactogenicity at the preclinical stage remains a major challenge in vaccine development, as conventional animal studies and in vitro assays capture general inflammation but fail to quantify local or systemic reactogenicity relevant to humans. Using transcriptomic data from the BioVacSafe consortium encompassing seven vaccines and immunostimulants in mice and five licensed vaccines in humans, we developed a cross-compartment and cross-species predictive model of vaccine reactogenicity. Reactogenicity classes were defined in mouse muscle based on the magnitude of transcriptomic responses and literature evidence. A penalized ordinal regression model was trained to predict both discrete classes and continuous scores of reactogenicity. Transcriptomic profiles from mouse muscle were highly predictive of reactogenicity, with key genes enriched in inflammatory and tissue repair pathways such as IL6/JAK/STAT3 signalling. The model retained strong performance when transferred to mouse blood and revealed shared transcriptional programs between compartments, suggesting coordinated innate responses. When applied to human blood, the classifier correctly ranked licensed vaccines by reactogenicity, identifying Fluad (MF59-adjuvanted) as the most reactogenic, in agreement with elevated C-reactive protein and ReactoScore values, while Engerix-B, Varilrix, and Stamaril were classified as low-reactogenicity formulations. These results align with clinical safety data and demonstrate that early transcriptomic signatures in mice can predict human reactogenicity profiles. Our study presents a pan-vaccine, cross-species transcriptomic signature that bridges preclinical and clinical data, offering a foundation for translational biomarkers and mechanism-informed assessment of vaccine tolerability.

Introduction

Vaccine reactogenicity refers to the local and systemic clinical manifestations that occur after vaccination, partly driven by inflammation resulting from the recruitment and activation of immune cells. Although most events are transient and mild, their frequency and severity vary across vaccine platforms (e.g., adjuvanted protein, viral vector, mRNA) and between individuals¹. This variability complicates both clinical assessment and regulatory decision-making².

The COVID-19 pandemic highlighted the need for reliable approaches to evaluate vaccine reactogenicity not only prior to clinical testing but also during vaccination campaigns, where responses have been shown to vary according to age, sex, and pre-existing comorbidities³. Across the vaccine development pipeline, reactogenicity and safety are monitored, beginning with *in vitro* assays and animal toxicology studies⁴, progressing through clinical trials, and extending to post-licensure pharmacovigilance². Preclinical evaluations are intended to exclude highly reactogenic candidates early, thereby reducing costs and focusing resources on the most promising formulations. Nevertheless, the mechanisms underlying local and systemic symptoms after vaccination remain poorly understood³, and no reliable biomarkers have yet been established to predict human reactogenicity². The discovery of translatable markers of reactogenicity therefore remains an active area of research.

To address these limitations, translational research increasingly relies on preclinical models and microphysiological platforms, including organs-on-chips and *in vitro* assays⁵, to elucidate the mechanisms of action of vaccines and evaluate their associated adverse events. For example, the Modular Immune *In vitro* Construct platform⁶ predicts local and systemic adverse events by profiling cytokine and chemokine release from PBMCs. Model performance is driven by four mediators (IL-1 β , IL-6, IL-10, and CCL4), and when applied to Trumenba, a vaccine not included in training, the model accurately ranked observed local and systemic reactogenicity. While representing the first quantitative framework for reactogenicity prediction, the current results remain constrained by the lack of paired biological and clinical data as well as its reliance on a limited panel of markers.

Preclinical studies in mice and rabbits, on the other hand, provide information on innate immune activation, cytokine release, and physiological endpoints, with rabbits serving as the regulatory standard for vaccine toxicity assessment⁴. These studies typically rely on markers such as IL-6, CRP, hepatic enzymes, and histopathology⁵. While informative, such endpoints mainly capture general inflammatory responses rather than specific features of vaccine reactogenicity. In comparative adjuvant studies in adults receiving hepatitis B antigen formulated with AS01B or Alum, elevated IL-6 and IFN- γ levels were associated with systemic reactogenicity following AS01B administration⁷. Similarly, findings from the BioVacSafe studies revealed conserved interferon and chemokine-related transcriptional modules such as IL-10 and CXCL10, that were induced in both mouse and human datasets^{8,9}. Yet, these cross-species commonalities remain largely descriptive and do not provide quantitative information on the magnitude or type (local versus systemic) of reactogenicity that may occur in humans. Overall, these studies underscore both the value and the limitations of early-phase biomarkers and highlight the need for improved translational models that bridge preclinical readouts with clinical tolerability outcomes.

The present study builds on BioVacSafe, an Innovative Medicines Initiative-funded public-private partnership aimed at systematically identifying biomarkers of relatively common inflammatory events after adjuvanted immunization using human, animal, and population-based models¹⁰. This consortium pioneered harmonized multi-tissue transcriptomics and correlation-based analyses to identify biomarkers of vaccine reactogenicity across species^{8,9}. Although these studies provided key mechanistic insights into vaccine reactogenicity, to our knowledge no predictive signature (*i.e.* set of genes) has yet been reported, either within the consortium or by other groups. To address this gap, we adopt a predictive framework aimed at identifying translational biomarkers of reactogenicity from animal to human. Based on BioVacSafe transcriptomic data (Figure 1a,b¹¹), we first derive a time-specific signature of reactogenicity from mouse muscle following exposure to four licensed vaccines with established safety profiles and three TLR agonists with defined inflammatory potentials (Figure 1d¹²). The diversity of these formulations provides a robust basis to capture genuine reactogenicity patterns and supports generalization to a broad range of vaccine candidates. We then tested this reactogenicity signature in mouse blood to assess its applicability in a more accessible compartment (Figure 1d¹³), facilitating future biomarker studies. Finally, we applied the reactogenicity signature to human blood across additional vaccine formulations to evaluate cross-species translatability. This three-stage strategy demonstrates that signatures identified in mouse muscle can also be detected in mouse and human blood, supporting the use of

animal data to predict vaccine safety outcomes in clinical settings. Of note, to train the signature in mouse muscle, reactogenicity classes are defined by combining the magnitude of the transcriptomic response in mice with evidence from the human literature. For simplicity, we therefore retain the term “reactogenicity” for both species, even though mouse models are not typically used to define it. Furthermore, the terms signature, classifier and model are used interchangeably in the rest of the paper.

Results

Defining Reactogenicity Classes from Transcriptomic and Literature Evidence

In the absence of head-to-head clinical reactogenicity data for the seven formulations, no foundational dataset was available to train the predictive model. We thus established a provisional hierarchy by using the first principal component of mouse muscle as proxy for inflammation, assuming that inflammatory response represented the main source of variation during the early post-vaccination period. This inferred hierarchy was then reconciled with published vaccine literature to define reactogenicity classes. These classes were subsequently used to train a penalized ordinal regression model on mouse muscle, the compartment in which vaccine response was found to be strongest in the original study⁸. The fitted model was then applied on the three transcriptomic datasets (mouse muscle and blood, human blood), to predict a continuous reactogenicity score, hereafter referred to as the *latent variable* (see **Materials and Methods**), as well as reactogenicity classes obtained by discretizing this latent variable.

In this way, the effects of time and vaccine formulation were examined along the first two principal components (PCs) in each dataset (Supplementary Figure 1 [↗](#)). Although the proportion of explained variance captured by these PCs was low in all four datasets, distinct patterns emerged. In mouse muscle, vaccine formulations were broadly distributed along the first principal axis in the following order: Saline, Agrippal, Engerix B, Poly I:C, Fluad, IFA, LPS and Pentavac SD. This gradient suggested that the formulations elicited qualitatively similar responses that differed mainly in magnitude. The second axis captured temporal variation, with values increasing across successive time points. Thus, both biological differences between formulations and temporal dynamics were detected, despite the relatively low explained variance. A comparable, though less pronounced, pattern was observed in mouse blood: LPS and Pentavac SD remained distinct from other vaccines; early (4–24h) and late (48–168h) time points exhibited differing directions. In the combined mouse dataset, the two compartments (muscle versus blood) were separated along a direction that was mostly aligned with the first principal component. Similarly, the vaccine formulations were distributed along a trajectory that followed approximately the second component, suggesting that tissue differences had a stronger effect than vaccine-induced variation. Of note, ellipses showed same orientation, suggesting that the vaccine effect was conserved across tissues. In contrast, the human dataset showed substantial overlap across both vaccine types and time points, pointing to a weaker signal.

When examined at individual time points, the two mouse datasets recapitulated a similar hierarchy with clearer separations than in the joint analysis (Supplementary Figure 2 [↗](#)). As already noted in the initial study⁸, this hierarchy was not strictly conserved over time due to distinct kinetics among formulations. Because our goal was to capture the overall hierarchy of reactogenicity rather than time point-specific differences, we derived a classification that remained broadly stable and homogeneous across time points. Owing to the limited availability of direct comparisons between formulations in the literature, we grouped them along the first principal component into three classes: low (Poly I:C, Agrippal, Engerix B and Saline); medium (IFA and Fluad), and high (LPS and Pentavac SD). These classes were clearly visible at 24 and 48 hours in mouse muscle (Supplementary Figure 2 [↗](#)). In contrast, in mouse blood only the high class remained separated from the low and medium groups, in line with the observations on aggregated time points (Supplementary Figure 1 [↗](#)).

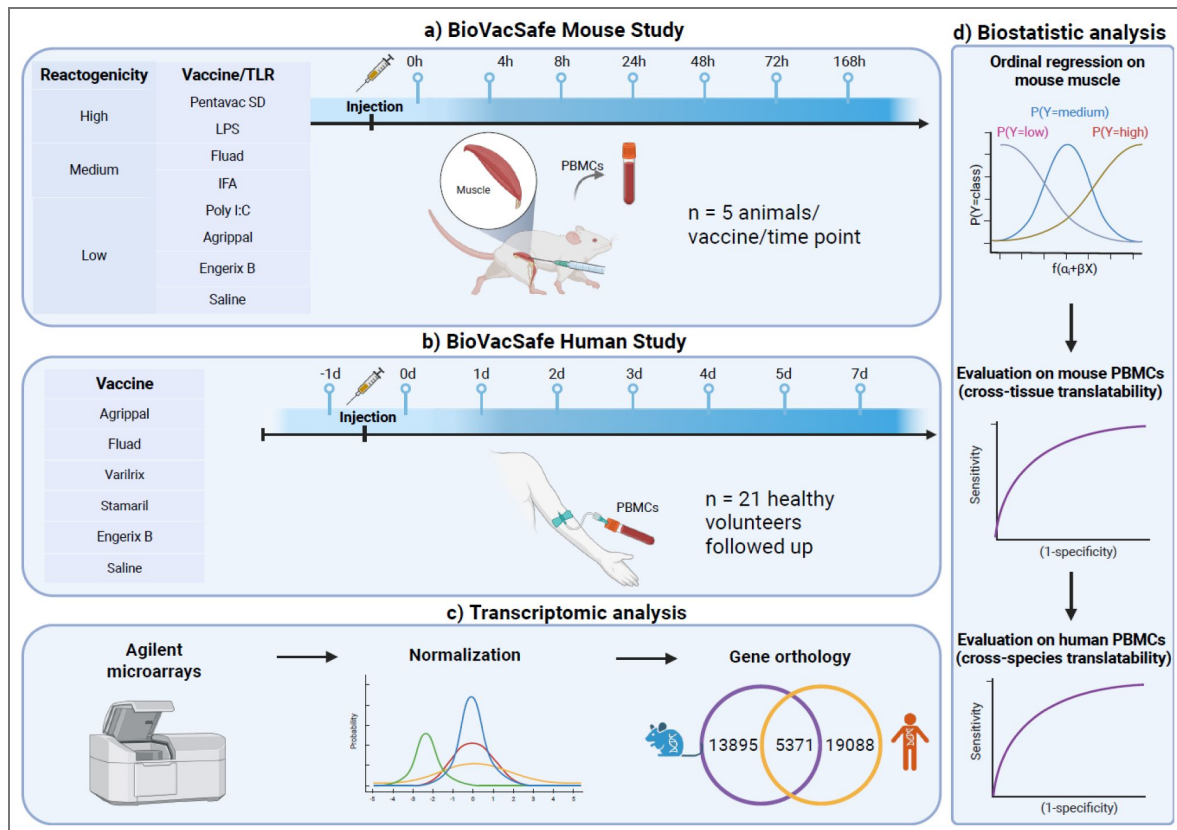


Figure 1. Study analysis workflow.

BioVacSafe (a) mouse and (b) human transcriptomic datasets were retrieved, (c) normalized, and mapped through orthologous probes to enable cross-species comparison. A penalized ordinal regression model was trained on mouse muscle data to derive a time-specific reactogenicity signature. (d) The signature was subsequently applied to mouse and human blood to assess its translatability across compartments and species.

Although reactogenicity levels of vaccine formulations varies considerably between humans and mice due to species-specific immune mechanisms and differences in vaccine/adjuvant properties¹¹, those three classes observed in mouse muscle are consistent with reactogenicity profiles described in the literature. LPS, a canonical inducer of systemic inflammation in mice, is among the most reactogenic agents¹². In humans, LPS toxicity has precluded direct use, leading to detoxified derivatives such as monophosphoryl lipid A, which retains immune stimulation while reducing harmful effects¹³. Reactogenicity in Pentavac SD is driven by its whole-cell pertussis component, which is inherently more reactogenic than acellular alternatives¹⁴.

Among the medium-reactogenicity formulations, Montanide (IFA-type) emulsions have been demonstrated to reliably boost immunogenicity and are characterized by dose- and formulation-dependent local reactogenicity such as commonly pain, nodules or granulomas^{15–17}. In clinical trials using Montanide with peptide antigens (e.g. HIV and malaria vaccines), sterile abscesses were reported, limiting their broader prophylactic use¹⁸. Compared with non-adjuvanted influenza vaccines, MF59-adjuvanted formulations were shown to have higher rates of mild-moderate reactions at local injection site but comparable AEs¹⁹.

Within the low-reactogenicity class, Poly I:C adjuvantation was investigated in human vaccine trials and glioblastoma settings, where it was described as safe and well tolerated, with predominantly low-grade local reactions (e.g. mild injection site effects) and no clear pattern of severe systemic toxicity^{20,21}. Across 36 clinical trials, Engerix-B displayed modest adverse event rates, with injection-site soreness reported in ~22% of recipients and fatigue in ~14%, confirming its overall favourable tolerability profile²².

Transcriptomic profiles in mouse muscle are highly predictive of vaccine reactogenicity

Building upon these reactogenicity classes, we trained and evaluated an ordinal regression model on mouse muscle. These two steps were applied to both the complete mouse microarray and the subset of probes corresponding to genes orthologous to human genes. The weighted F1 score, combining precision and recall across classes, indicated that all six time points were predictive of reactogenicity class, with peak performance observed at 8h and 24h in the full set (median weighted F1 = 0.98; Figure 2a [↗](#)). This was consistent with the greater class separation seen at 24h in the PCA (Supplementary Figure 2 [↗](#)). Interestingly, the median F1 scores were identical between the two probe sets across all time points except 8h, suggesting that the reactogenicity signal was largely preserved in orthologous genes. To contextualize these results in terms of classification error, confusion matrices were generated. Across 100 cross-validation repetitions, the average number of misclassified samples, out of 40 samples per time point, ranged from one at 24h to more than three at 4h and 168h (Supplementary Figure 3a [↗](#)), corroborating the higher predictive performance at 24h.

We next sought to characterize the main drivers of the reactogenicity signature, i.e., genes with large selection frequency (Supplementary Table 1). We first conducted a differential analysis between high and low classes to quantify effect size and assess statistical significance (Supplementary Table 2). Although $\log_2$ fold-change did not correlate globally with selection frequency, reflecting the distinct statistical properties of univariate and multivariate analyses, the comparison revealed that the most predictive genes (frequency > 10%) almost systematically exhibited $|\log_2 \text{fold-change}| > 1$ and adjusted p-values < 10^{-5} , thresholds commonly considered biologically meaningful (Supplementary Figure 3b [↗](#)). Visual inspection of the top 10 most frequently selected genes at each time point (Figure 2b [↗](#)) confirmed that these 27 unique genes underwent substantial modulation in one or more reactogenicity classes, with $\log_2$ fold-changes reaching up to 2 (Figure 2c [↗](#)). The relative ordering of classes was preserved across all genes, indicating that the ordinal regression model effectively identified genes associated with graded vaccine reactogenicity. All genes showed maximal class separation between 24 and 72 hours, consistent with the classification performance trends described above.

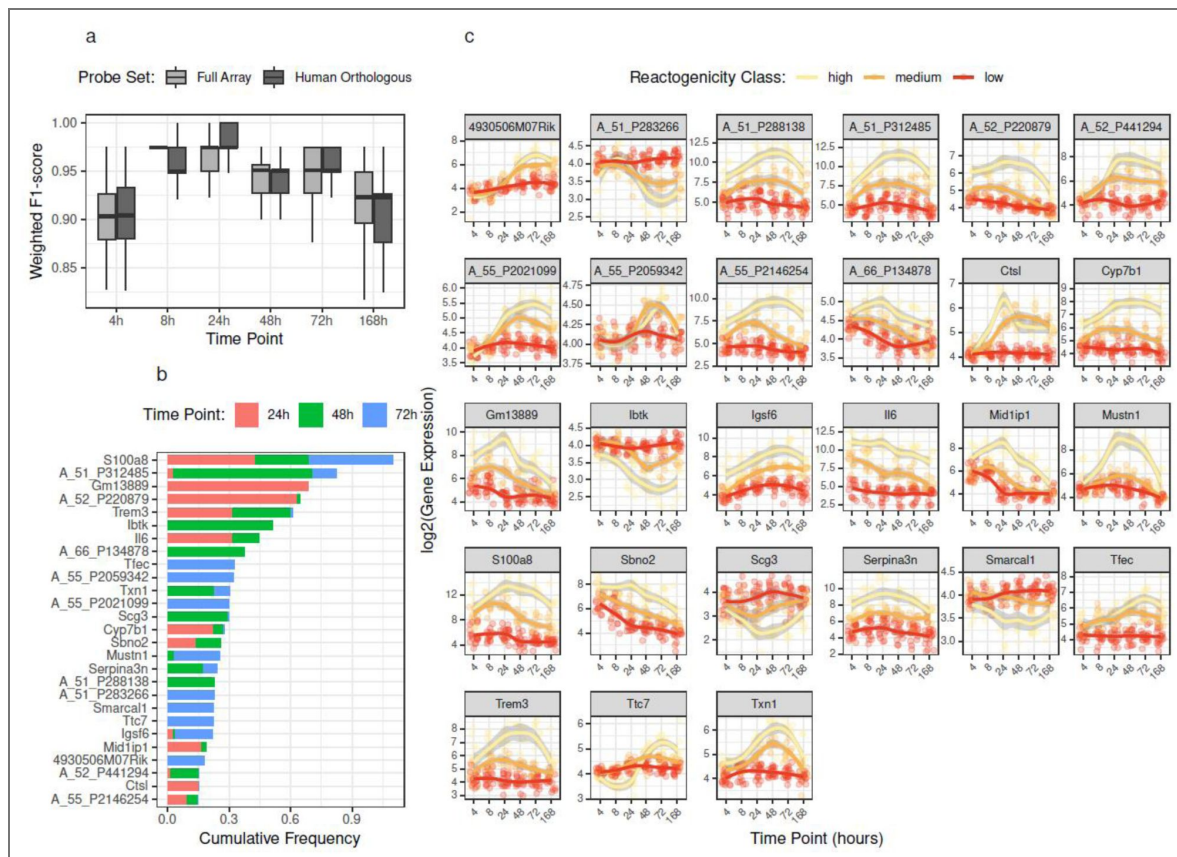


Figure 2. Pan-vaccine signature definition and evaluation in mouse muscle.

(a) A penalized ordinal regression model was trained and evaluated using repeated nested cross-validation at each time point, with performance assessed by the weighted F1-score on both the full mouse microarray and the subset of probes orthologous to human genes. **(b)** Gene importance for predicting reactogenicity was assessed by stability selection. The top 10 most frequently selected genes at 24, 48, and 72 hours are displayed. **(c)** Expression trajectories of the top 27 genes are shown over time for each reactogenicity class.

Additionally, we performed a pathway enrichment analysis on the selection frequency at each time point. Although none of MSigDB Hallmark gene sets reached statistical significance (p-values > 0.12), the seven pathways highlighted by GSEA aligned well with inflammatory processes such as TNF α Signalling via NF- κ B, Inflammatory Response and IL-6/JAK/STAT3 signalling (Supplementary Figure 3c [↗](#)). Furthermore, IL-6, which was among the top 10 most selected gene at 24h, has been reported to be a key component of vaccine-induced reactogenicity in the literature^{6,7,23}. Epithelial–Mesenchymal Transition was also highlighted by pathway enrichment, and is known to play a role in tissue repair²⁴, suggesting that reactogenicity prediction does not solely rely on inflammatory markers, reflecting the general significance of inflammation and of acute phase reactants²⁵.

The reactogenicity signature translates across compartments

Following the identification of a pan-vaccine classifier in mouse muscle, we applied the model to mouse blood to evaluate its translatability across compartments. This step was motivated by the predominance of blood-based immune monitoring in vaccine studies and by the practical consideration that any clinically deployable signature would need to be measurable in peripheral blood, where sampling is far less invasive than muscle biopsies.

The median weighted F1 ranged from 0.57 to 0.80, *i.e.* up to 0.35 lower than the corresponding performance in muscle at 24h and 48h (Supplementary Figure 4a [↗](#)). Despite this decrease, examination of Figure 3a [↗](#) revealed that Pentavac SD and LPS consistently exhibited higher median latent variables (*i.e.* continuous reactogenicity scores) than the other six vaccines across time-point combinations, with the most pronounced differences observed when applying the model at 72h in blood (see pairwise t-test in Supplementary Figure 4d [↗](#)). By contrast, the medium class showed minimal separation from the low-reactogenicity formulations, a finding consistent with PCA trends. These results indicate that the transition from muscle to blood primarily impact the distinction between low and medium reactogenicity groups, a pattern also visible in the dynamics of gene expression (Figure 3c [↗](#)). Given that the intermediate reactogenicity class likely reduces model performance, we retrained the model in muscle using low and high reactogenicity classes and reapplied it to blood data. This refined model achieved better separation, with a median weighted F1 = 0.95 across all time points (Supplementary Figure 4a [↗](#)), pointing to robust translatability of the model to mouse blood when the intermediate class was excluded. Of note, classification was maximal (weighted F1 = 1.00) at 72 h in muscle (Supplementary Figure 4c [↗](#)), the time point at which the log₂ fold-change between low and high reactogenicity across compartments correlated most strongly (R = 0.59, Supplementary Figure 4b [↗](#)). These higher classification and correlation values found at 72 h suggest that, despite differences in the magnitude and kinetics of transcriptomic profiles across muscle, blood, and lymph nodes⁸, the responses in muscle and blood exhibit the greatest similarity at this time point. These observations point toward shared transcriptional programs between compartments, potentially driven by immune cells trafficking from the injection site to the circulation.

The reactogenicity signature translates across species and is consistent with inflammation readouts

We finally investigated whether the pan-vaccine classifier generalize across species by applying it to human blood using both two-class and three-class models. In the absence of comparative clinical studies directly assessing vaccine reactogenicity side by side, as noted earlier, and given that PCA revealed no clear separation between formulations in human blood, the definition of reactogenicity classes was consequently not performed in humans. Instead, we assessed the consistency between our model predictions and inflammation readouts (ReactoScore and CRP).

Despite smaller variability across formulations than in mouse blood, Flud displayed higher values than the other vaccines in CRP, ReactoScore, and latent variables predicted from mouse muscle at 24h (Figure 4 [↗](#)). These differences were significant for both inflammation readouts and for the two-class model applied to human blood at 24h (Supplementary Figure 5 [↗](#)). By contrast, only CRP and ReactoScore exhibited higher levels for Agrippal and Engerix-B than for Saline,

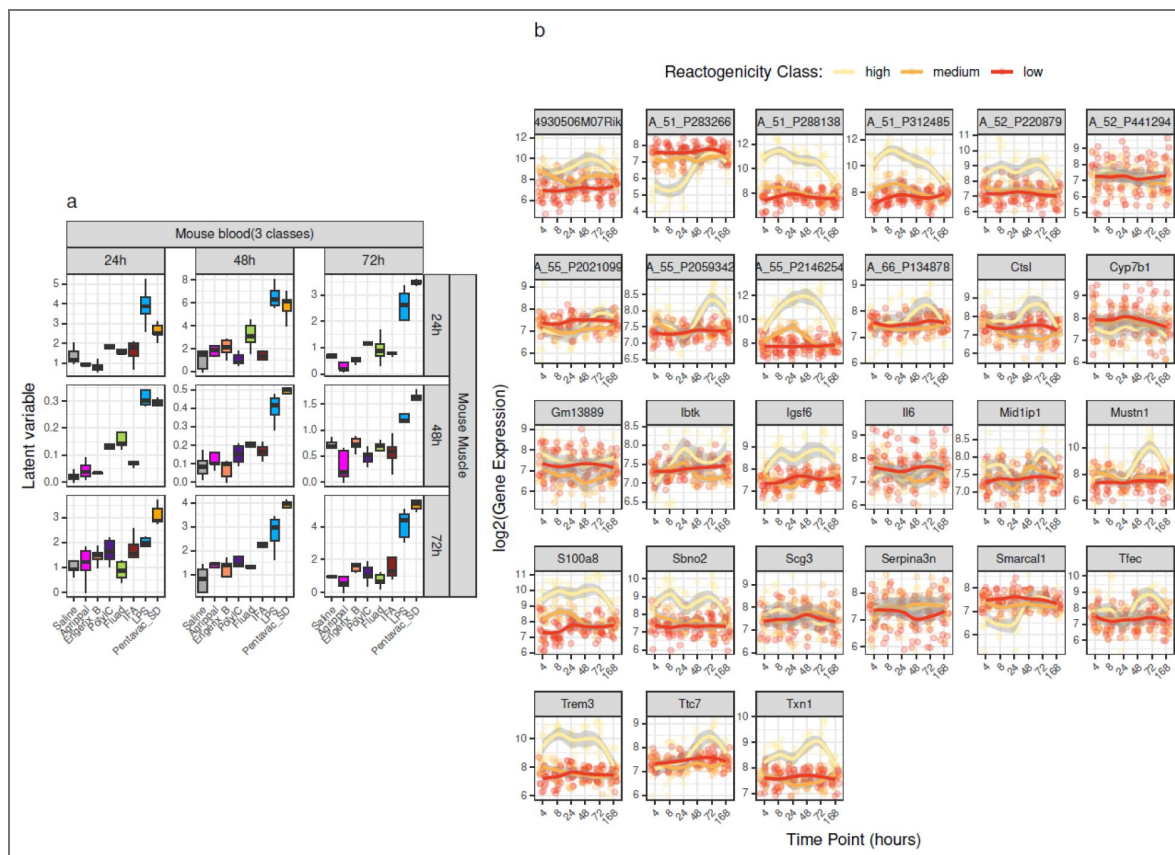


Figure 3. Predicted reactivity in mouse blood.

(a) Latent variables predicted across all combinations of three time points (24h, 48h, 72h), with models trained in mouse muscle and applied in mouse blood. **(b)** Temporal expression profiles of the most predictive genes for each reactivity class in mouse blood.

Varilrix, and Stamaril across all time points, a pattern not captured by the model. Despite the absence of gold standard, the increased predicted reactogenicity in Fludac is consistent with reported safety profiles. As mentioned earlier, in a meta-analysis of 24 studies, Domnich *et al.* concluded that MF59-adjuvanted flu vaccines are generally more immunogenic, induce more reactogenic events, without increasing serious adverse events¹⁹. Stamaril^{26,27}, Varilrix and Engerix B²⁹ on the other hand are widely regarded as having good safety profiles. Taken together, these findings show that the predicted reactogenicity aligns with established clinical safety data, supporting the ability of the transcriptomic signature to distinguish between reactogenicity levels across vaccine formulations. This underscores its potential as a preclinical tool for anticipating the inflammatory profile of novel vaccine candidates.

Discussion

Despite their central role in vaccine safety, the cellular and molecular mechanisms of vaccine reactogenicity remain incompletely understood. Upon vaccination, clinical signs related to reactogenicity are observed both locally and systemically. Clinical symptoms are hard to predict prior to injection, because they are a product of various direct and indirect mechanisms that elicit different reactions². A major outlying question in the field is how diverse molecular mediators that are detected at different time points post vaccination translate into clinical events.

Preclinical assessment of vaccine reactogenicity is therefore critical to ensure safety before first-in-human trials. Yet current practice relies on a narrow set of surrogate markers, such as hepatic enzymes for toxicity or CRP in rabbits for inflammation³⁰. The rapid deployment of mRNA vaccines during the COVID-19 pandemic further exposed the limitations of these tools, underscoring the need for predictive, mechanism-informed biomarkers that can be integrated early in vaccine development^{10,31}.

To overcome this limitation, we developed a pan-vaccine transcriptomic signature of reactogenicity by leveraging data from the BioVacSafe consortium. We first focused on the injection site, where transcriptional changes are stronger than in peripheral blood⁸ and where diverse immune cells are rapidly recruited upon vaccination^{32–34}. Licensed vaccines and immunostimulants were classified into three reactogenicity categories based on PCA results and prior knowledge of their inflammatory profiles. From mouse muscle, we identified a predictive signature that remained robust when transferred to blood after excluding the intermediate reactogenicity class. When applied to human data, the classifier correctly identified Fludac as the most reactogenic vaccine, consistent with CRP levels, ReactoScore, and established clinical safety profiles. The wide diversity of formulations used in mice, from saline to LPS, provides a unique advantage: the model was trained on across a reactogenicity spectrum wider than that observed among licensed human vaccines, enabling it to predict whether a new formulation lies within a “tolerability zone” or exceeds it. Collectively, these findings demonstrate the potential of a cross-compartment, cross-species framework, agnostic to vaccine platform.

Although marked differences exist between mice and humans in pattern-recognition receptor expression, signalling pathways, immune cell subset composition, cytokine responses and regulatory networks³⁵, the strong performance of the signature across datasets, together with the diversity of formulations in the training set, suggests that the model captures conserved host-response mechanisms across species. These results also support the notion that key components of vaccine-induced inflammation and reactogenicity are evolutionarily conserved. Viewing reactogenicity as part of a homeostatic process³⁶ may explain why predictive patterns of innate activation are preserved across species^{37,38}. In summary, when applied to licensed vaccines with well-established safety profiles in human, our signature aligns with standard CRP measurements and can be implemented in mouse models to infer vaccine safety at the preclinical stage. Importantly, unlike CRP which increases only moderately in mice compared to humans³⁹ and whose complement-activating role is not conserved across species^{39,40}, our transcriptomic signature turned out to be translatable.

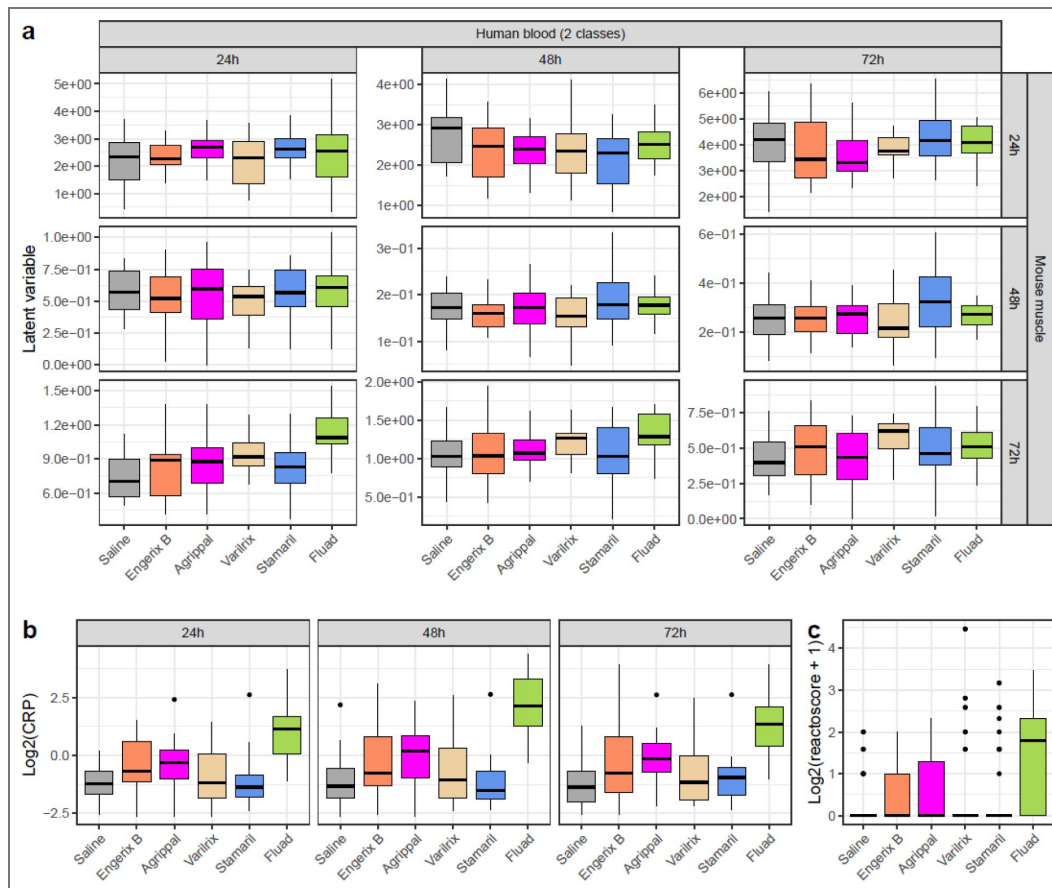


Figure 4. Predicted reactivity in human blood.

(a) Latent variables predicted across all combinations of three time points (24h, 48h, 72h), using models trained in mouse muscle and applied in human blood. (b) CRP and (c) ReactoScore measured in the same individuals for comparison.

Similar studies have reported predictive transcriptomic signatures of vaccine outcomes. For immunogenicity, both baseline⁴¹ and post-vaccination⁴² markers have been shown to effectively predict antibody responses, the former being derived from thirteen vaccines. For reactogenicity, an infant study of 4CMenB identified four post-vaccination markers of febrile responses in plasma proteins and was subsequently validated in mice⁴³. Our work extends the latter by providing a pan-vaccine, cross-species signature of reactogenicity based on early post-vaccination responses, demonstrating translational potential in preclinical safety evaluation.

As a next step, we envision validating the current signature across a broader panel of vaccine formulations and in human cell-based models to reinforce its functional relevance and predictive value⁴⁴. An additional line of development will involve assessing its applicability across emerging vaccine platforms, including viral vectors, mRNA formulations, and novel adjuvants such as polyacrylate⁴⁵, thereby evaluating its robustness in next-generation vaccine development. A subsequent signature reduction step, retaining only the most informative transcripts, will further enable its translation to low-throughput platforms such as multiplex qPCR. Ultimately, the signature could also be refined for specific populations, such as paediatric or immunocompromised individuals, enabling more accurate and clinically relevant assessment of vaccine reactogenicity.

Material & methods

BioVacSafe datasets

As described in the original study⁸, seven vaccines and immunostimulants were evaluated in the mouse study: Pentavac SD (Diphtheria, tetanus, pertussis (whole cell), hepatitis B (rDNA) and haemophilus type b conjugate vaccine), Agrippal (Trivalent flu subunits – H3N2, H1N1 and influenza B), Fluad (Trivalent flu subunits – H3N2, H1N1 and influenza B + MF59), Engerix B (Recombinant hepatitis B sAg absorbed on alum), IFA (Montanide ISA 51 VG), LPS (LPS-EB Ultrapure), and the TLR3 agonist Poly I:C (Polyinosinic:polycytidylic acid), alongside a saline control. Mice were culled at six time points post-injection (4h, 8h, 24h, 48h, 72h, and 168h). The injected muscle site and lymph nodes draining the hind leg quadriceps muscles were harvested. Peripheral blood was sampled from the mouse tail vein. Total RNA was isolated, labelled, hybridized and profiled on an Agilent microarray. Each vaccine group and time point included five animals.

The human study⁹ assessed five licensed vaccines, Agrippal, Fluad, Engerix B, Stamaril (live-attenuated YFV vaccine), Varilrix (live-attenuated varicella zoster virus) along with a saline control, the first three overlapping with those tested mice. Participants were followed longitudinally, peripheral blood was collected from one day before vaccination through 28 days post-vaccination. Only samples obtained up to seven days post-vaccination (D-1, D0, D1, D2, D3, D4, D5, D7) were included to match the time points in the mouse study. Transcriptomic profiling was performed on blood-derived RNA and inflammatory mediators, including C-reactive protein (CRP), were quantified. Adverse events (AEs) were systematically documented, including description (e.g. headache, pain at injection site), classification (e.g. nervous system disorders, general disorders and administration site conditions), severity grading (mild, moderate, or severe), and duration (in days). AEs deemed non-specific (e.g. fatigue, nausea, nasal congestion, sleep disorders etc), unrelated to vaccination, recorded after 7 days post-vaccination, associated with indwelling cannula, or arising from laboratory artifacts were excluded from the analysis. The list of AEs retained in the analysis is available Supplementary Table 3. Similar to the original study⁹, a ReactoScore was calculated for each patient as follows:

$$R_{ind} = \sum_{AE, ind} S_{AE} * D_{AE}$$

where S_{AE} is the AE severity (1 = mild, 2 = moderate, 3 = severe), and D_{AE} its duration in days. Of note, patients with no documented AE were assigned a ReactoScore of 0.

Data preprocessing

All data were retrieved from the NCBI Gene Expression Omnibus (GEO) portal^{46,47}, using the R package *GEOquery* (accession numbers: GSE120661 and GSE124533). Microarray data were preprocessed with the R package *limma*, using background correction (*backgroundCorrect*, normexp method) followed by quantile normalization across arrays (*normalizeBetweenArrays*). Control probes were removed, as well as probes with no corresponding gene symbol, Entrez gene ID or ENSEMBL ID. Moreover, we excluded probes of non-protein coding transcripts. In case of probe replicates, expression was averaged (function *avereps*).

Orthology between mouse and human Agilent microarrays

GPL annotation files for the mouse and human Agilent microarrays were retrieved from GEO. Probe coordinates were based on the MGCv37⁴⁸ (mouse) and GRCh37⁴⁹ (human) genome assemblies. Mouse probe positions were converted to GRCh37 using UCSC LiftOver⁵⁰. Both the converted mouse probes and the original human probes were then mapped to exons from the GRCh37 GTF. A mouse–human probe pair was defined as orthologous if both mapped to at least one common exon. After excluding one-to-many matches, 5,371 orthologous probe pairs were retained.

Principal component analysis

Principal Component Analysis (PCA) was performed separately on each dataset, as well as jointly on the mouse muscle and blood data, to enable comparison of vaccine effects across compartments. To make the comparison possible, the microarrays were quantile normalized together. For each dataset, sample projections were coloured by vaccine or time points and ellipses associated with reactogenicity classes were added.

Differential analysis

Transcriptomic data from mouse muscle were analyzed for differential expression using the *lmFit* function of the R package *limma*. As in the original study, models including all interactions between treatments and time points were fitted. Instead of constructing contrasts comparing each vaccine against pre-vaccination and saline controls at each time point, comparisons were made at the class level. Specifically, we contrasted the high-reactogenicity class (LPS and Pentavac SD) with the low-reactogenicity class (Engerix B, Agrippal, Poly I:C, and Saline) using double contrasts $(H_t - H_{0h}) - (L_t - L_{0h})$, where H denotes the high-reactogenicity group, L the low-reactogenicity group, and t the considered time point. P-values were adjusted for multiple testing using Benjamini–Hochberg correction⁵¹.

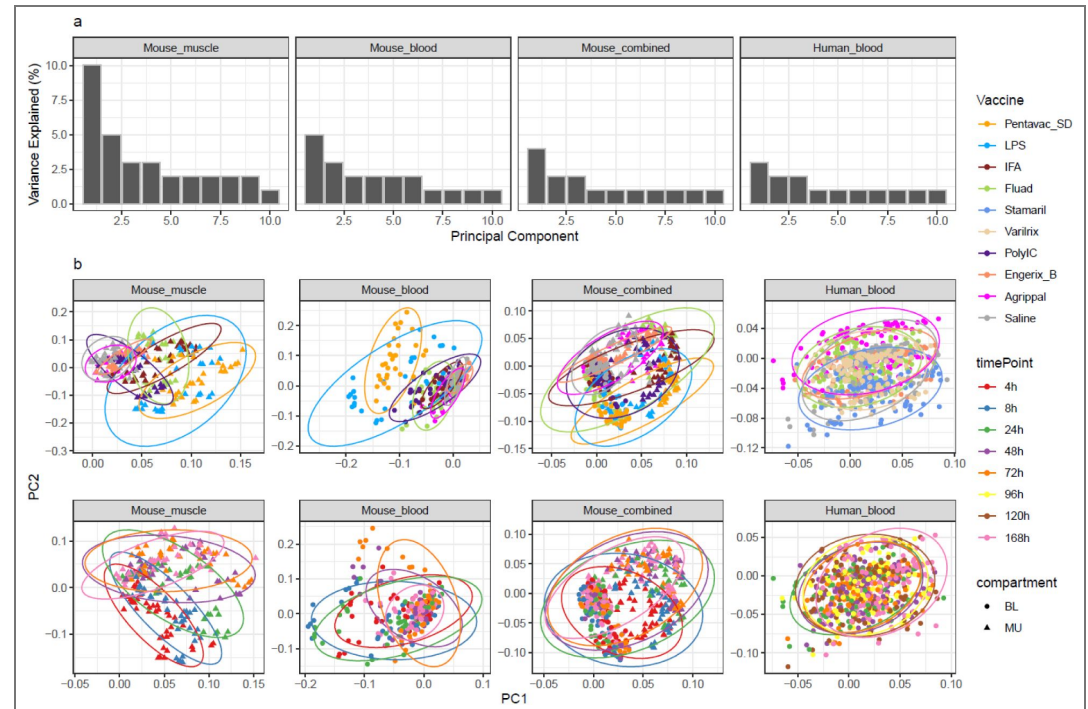
Predictive model

To predict reactogenicity from gene expression data, a penalized ordinal regression (*glmnetcr* R package) was applied to account for the ordered nature of the reactogenicity classes (Figure 1d [\[link\]](#)). The model incorporates an L1 penalty to enforce sparsity and enable variable selection. A key feature of ordinal regression is that it models the response variable as a discretized version of an unobserved continuous latent variable which we use as a continuous proxy for reactogenicity. The cut-points defining the class boundaries are ordered in a strictly increasing manner and are estimated along the regression coefficients. The model was first trained and evaluated on mouse muscle data, where coefficients and penalization level were estimated using 5-folds nested cross-validation. The cross-validation procedure was repeated 100 times to obtain a stable estimate of model performance and to quantify uncertainty in the performance metrics. To address class imbalance, observation weights were applied so that each class contributed equally to parameter estimation. The *Caret* R package was further used to generate balanced folds. The estimated coefficients were then applied on mouse and human blood data to compute latent variables and reactogenicity predictions. Model performance was evaluated using the weighted F1 score, which balances precision and recall across classes while accounting for their relative frequencies,

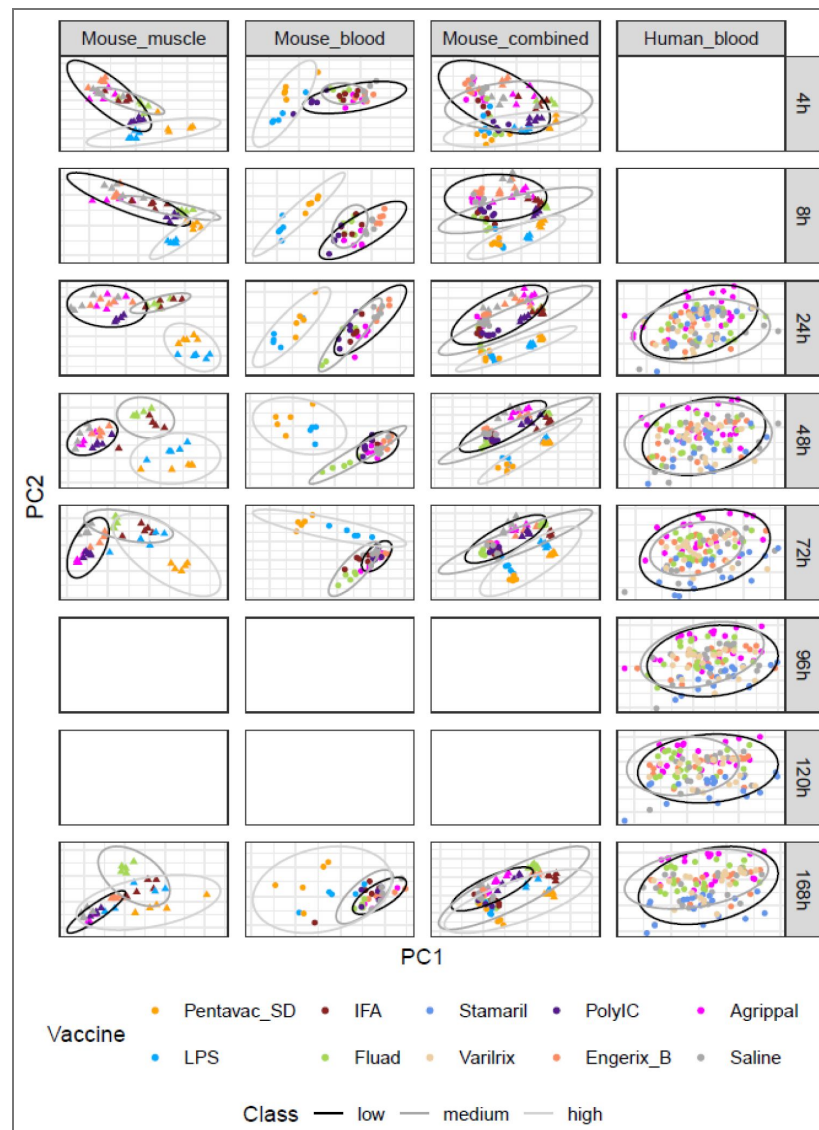
providing a robust measure in the presence of class imbalance. Weighted F1 score ranges from 0 to 1, where values close to 0 indicate poor classification performance while values approaching 1 indicate high precision and recall across all classes, reflecting excellent predictive performance. To identify the support of the molecular signature, variable importance was measured using stability selection⁵², a method based on subsampling.

To functionally characterize the signature, we performed a pathway enrichment analysis on gene selection frequencies using GSEA⁵³, assessing the association between gene ranking and MSigDB Hallmark gene sets.

Supplementary figures

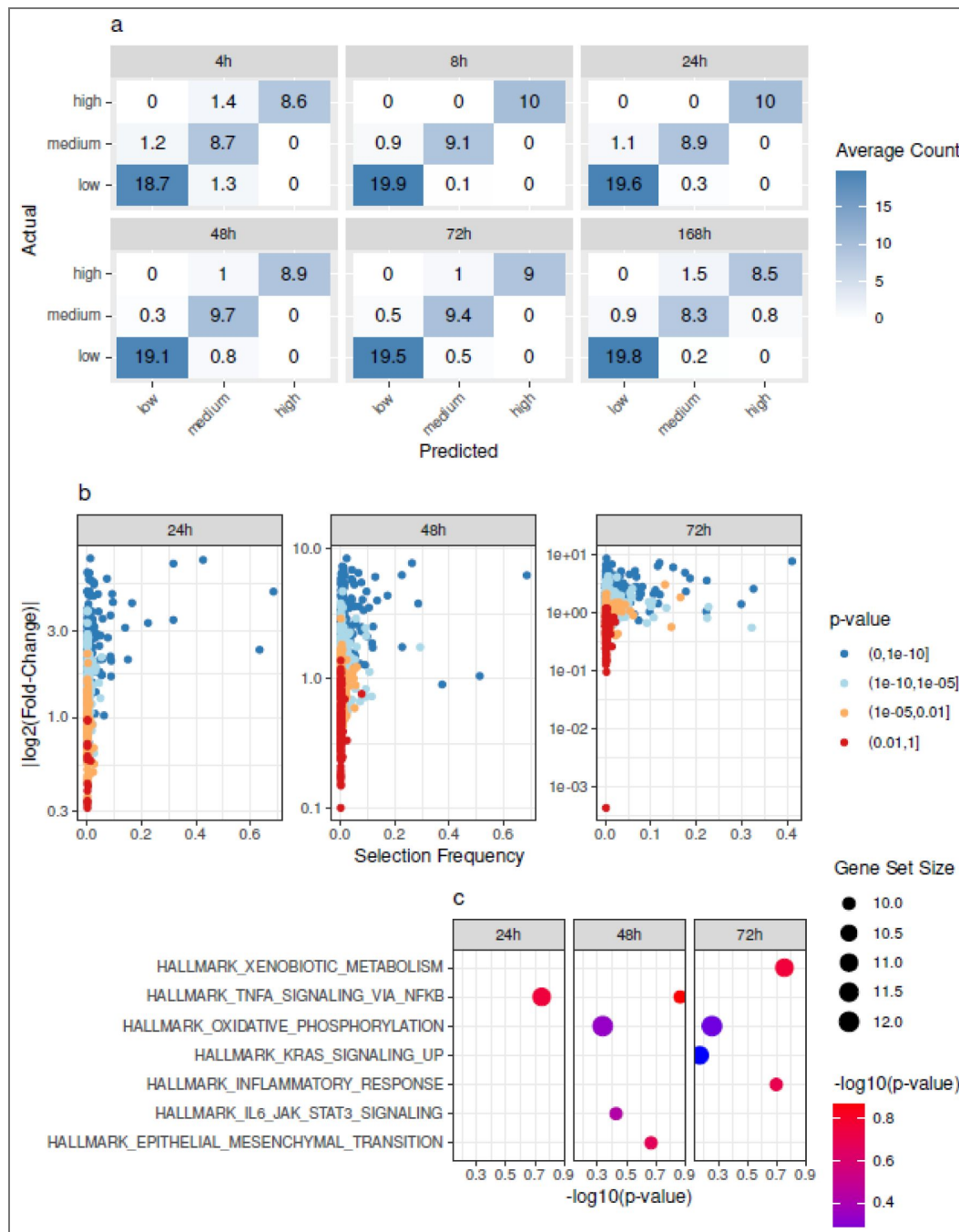


Supplementary Figure 1. Principal component analysis of vaccine formulations across three transcriptomic datasets. (a) Percentage of variance explained by the principal components. **(b)** Sample projections onto the first two principal components, coloured by vaccine formulation (top) and by time point (bottom).



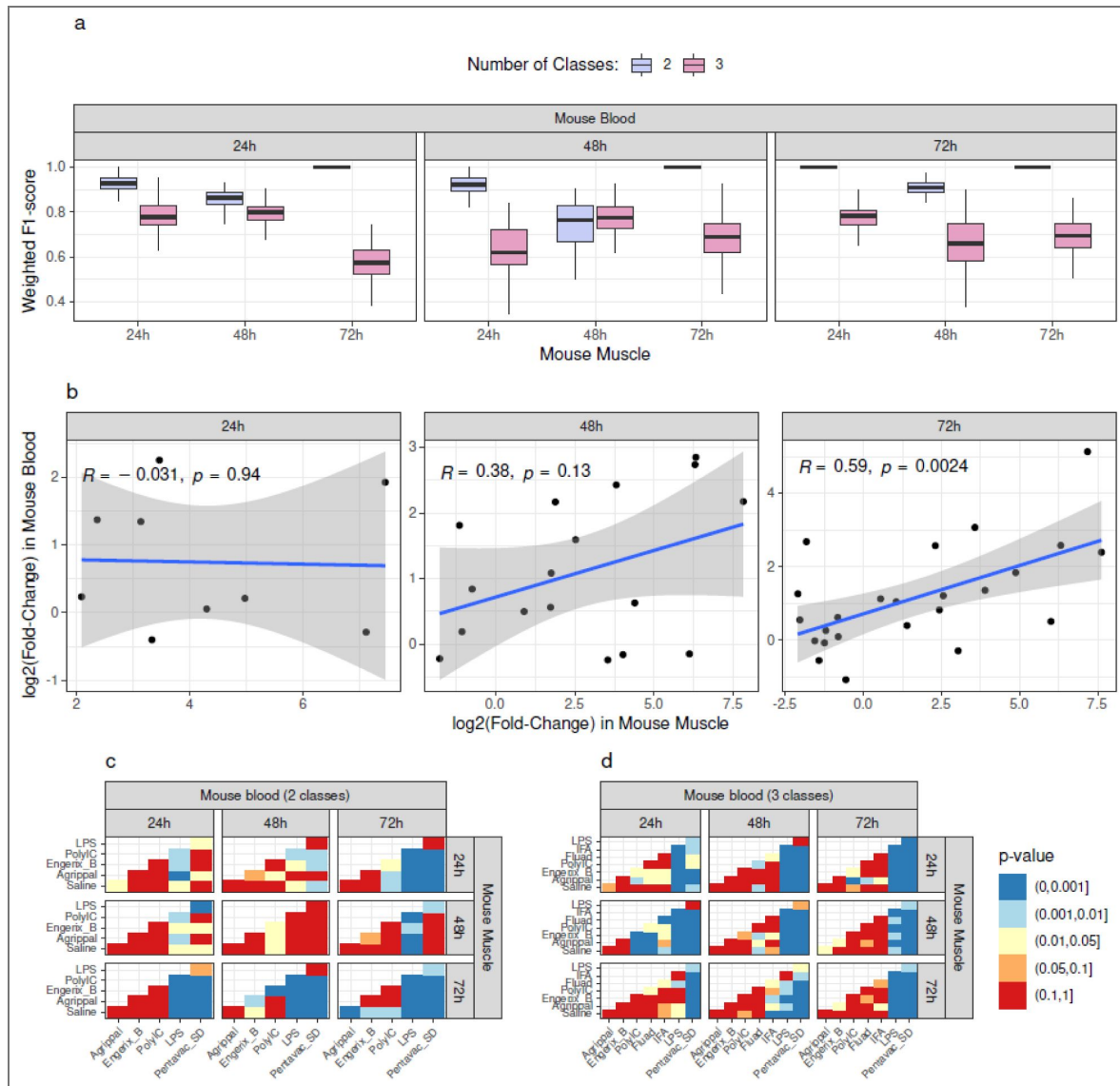
Supplementary Figure 2. Principal component analysis of vaccine formulations by time point across transcriptomic datasets.

Samples are projected onto the first two principal components, coloured by vaccine formulation, with ellipses representing reactivity classes.



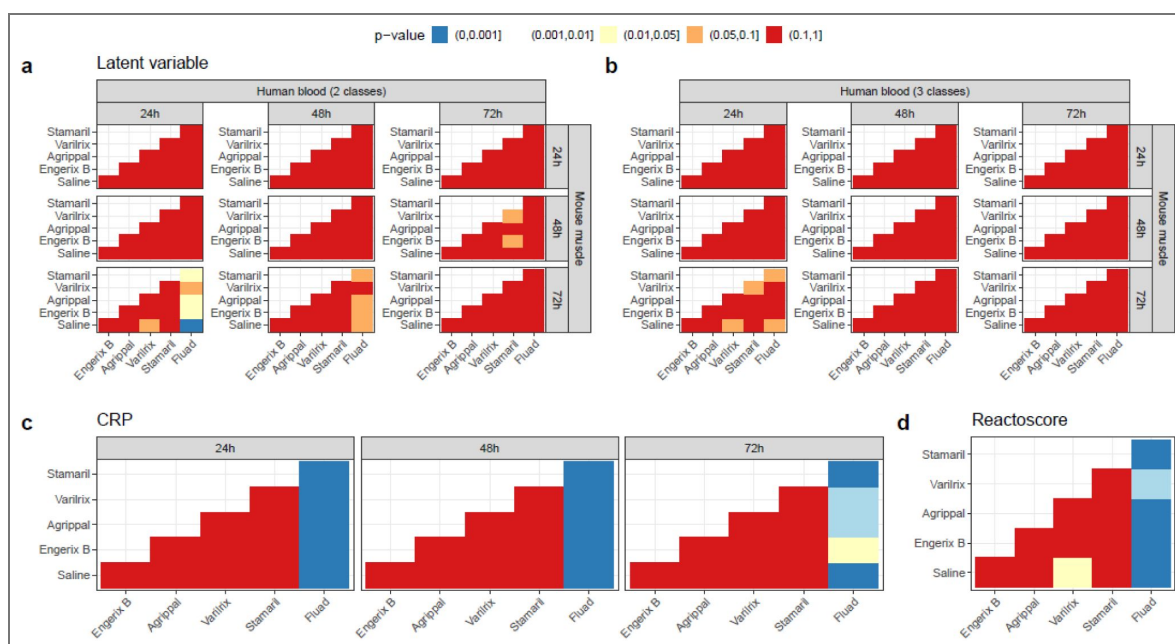
Supplementary Figure 3.

(a) Confusion matrices comparing predicted (columns) and observed (rows) reactogenicity classes in mouse muscle by time point. Results are averaged over 100 repetitions of nested cross-validation. **(b)** Comparison of gene predictive power (selection frequency computed via stability selection) and univariate differential expression (log fold change and p-value) between low and high reactogenicity classes at 24h, 48h and 72h. **(c)** Gene-set enrichment analysis on selection frequency, measuring the association between gene ranking and MSigDB gene sets.



Supplementary Figure 4.

(a) Once trained on mouse muscle, the two and three-class models were evaluated on mouse blood at all combinations of time points using the weighted F1-score. (b) Scatter plots showing the relationship between $\log_2$ fold-changes in mouse muscle and blood across low and high reactivity classes at 24h, 48h, and 72 h for the most predictive genes (selection frequency greater than 10%). (c) Adjusted t -test p -values for all pairwise comparisons of latent variables predicted at three time points (24h, 48h, and 72 h) using two and three-class models trained on mouse muscle and applied to mouse blood.



Supplementary Figure 5. Adjusted *t*-test *p*-values for all pairwise comparisons of (a) latent variables predicted at three time points (24h, 48h, and 72 h) using two and three-class models trained on mouse muscle and applied to human blood; (b) CRP and (c) reactoScore values.

Acknowledgements

We gratefully acknowledge Prof. Giuseppe Del Giudice and Prof. Alberto Mantovani for their input regarding the BioVacSafe datasets and helpful discussions on vaccine reactogenicity.

Additional information

Code availability

The code to reproduce the results is available on GitHub (<https://github.com/bioaster/>)

Author ORCID iDs

Laurent Beloeil: <https://orcid.org/0000-0003-3313-602X>

Additional files

Supplementary Table 1. [Variable selection frequency computed with stability selection on mouse muscle at 24h, 48, and 72h.](#)

Supplementary Table 2. [Differential analysis between low and high reactogenicity classes in mouse muscle.](#)

Supplementary Table 3. [Adverse event classification.](#)

References

1. Chapin-Bardales J., Gee J., Myers T (2021) Reactogenicity Following Receipt of mRNA-Based COVID-19 Vaccines. *JAMA* **325**:2201-2202
2. Hervé C., Laupèze B., Del Giudice G., Didierlaurent A. M., Tavares Da Silva F. (2019) The how's and what's of vaccine reactogenicity. *npj Vaccines* **4**:39
3. Holt H., et al. (2023) Incidence determinants and serological correlates of reactive symptoms following SARS-CoV-2 vaccination. *npj Vaccines* **8**:26
4. Di Pasquale A., et al. (2016) Vaccine safety evaluation: Practical aspects in assessing benefits and risks. *Vaccine* **34**:6672-6680
5. Zaitseva M., et al. (2012) Use of human MonoMac6 cells for development of in vitro assay predictive of adjuvant safety in vivo. *Vaccine* **30**:4859-4865
6. Pullen R. H., et al. (2024) A Predictive Model of Vaccine Reactogenicity Using Data from an In Vitro Human Innate Immunity Assay System. *The Journal of Immunology* **212**:904-916
7. Burny W., et al. (2019) Inflammatory parameters associated with systemic reactogenicity following vaccination with adjuvanted hepatitis B vaccines in humans. *Vaccine* **37**:2004-2015
8. McKay P. F., et al. (2019) Identification of potential biomarkers of vaccine inflammation in mice. *eLife* **8**:e46149 <https://doi.org/10.7554/eLife.46149>
9. Weiner J., et al. (2019) Characterization of potential biomarkers of reactogenicity of licensed antiviral vaccines: randomized controlled clinical trials conducted by the BIOVACSAFE consortium. *Sci Rep* **9**:20362
10. Lewis D. J. M., Lythgoe M. P (2015) Application of "Systems Vaccinology" to Evaluate Inflammation and Reactogenicity of Adjuvanted Preventative Vaccines. *Journal of Immunology Research* 1-11
11. Aucouturier J., Dupuis L., Deville S., Ascarateil S., Ganne V (2002) Montanide ISA 720 and 51: a new generation of water in oil emulsions as adjuvants for human vaccines. *Expert Rev Vaccines* **1**:111-118
12. Zhang X., et al. (2024) Application of lipopolysaccharide in establishing inflammatory models. *International Journal of Biological Macromolecules* **279**
13. Johnson A. G., Tomai M., Solem L., Beck L., Ribi E (1987) Characterization of a nontoxic monophosphoryl lipid A. *Rev Infect Dis* **9**:S512-516

14. **Szweijser-Zawislak E., et al. (2023)** Evaluation of Whole-Cell and Acellular Pertussis Vaccines in the Context of Long-Term Herd Immunity. *Vaccines* **11**
15. **Langermans J. A. M., et al. (2005)** Effect of adjuvant on reactogenicity and long-term immunogenicity of the malaria Vaccine ICC-1132 in macaques. *Vaccine* **23**:4935-4943
16. **Pierce M. A., et al. (2010)** Phase 1 Safety and Immunogenicity Trial of the Plasmodium falciparum Blood-Stage Malaria Vaccine AMA1-C1/ISA 720 in Australian Adults. *Vaccine* **28**:2236-2242
17. **Graham B. S., et al. (2010)** Immunization with Cocktail of HIV-Derived Peptides in Montanide ISA-51 Is Immunogenic, but Causes Sterile Abscesses and Unacceptable Reactogenicity. *PLoS One* **5**
18. **van Doorn E., Liu H., Huckriede A., Hak E (2015)** Safety and tolerability evaluation of the use of Montanide ISATM51 as vaccine adjuvant: A systematic review. *Hum Vaccin Immunother* **12**:159-169
19. **Domnich A., Trombetta C.-S., Fallani E., Salvatore M (2024)** Immunogenicity and safety of the MF59-adjuvanted seasonal influenza vaccine in non-elderly adults: A systematic review and meta-analysis. *PLOS One* **19**
20. **Levy Y., et al. (2024)** Safety and immunogenicity of CD40.HIVRI.Env, a dendritic cell-based HIV vaccine, in healthy HIV-uninfected adults: a first-in-human randomized, placebo-controlled, dose-escalation study (ANRS VRI06). *eClinicalMedicine* **77**:102845
21. **De Waele J., et al. (2021)** A systematic review on poly(I:C) and poly-ICLC in glioblastoma: adjuvants coordinating the unlocking of immunotherapy. *J Exp Clin Cancer Res* **40**
22. **GlaxoSmithKline Biologicals (2023)** ENGERIX-B. US Food and Drug Administration. <https://www.fda.gov/vaccines-blood-biologics/vaccines/engerix-b>
23. **Sheerin D., et al. (2019)** Comparative transcriptomics between species attributes reactogenicity pathways induced by the capsular group B meningococcal vaccine, 4CMenB, to the membrane-bound endotoxin of its outer membrane vesicle component. *Scientific Reports* **9**, 13797
24. **Youssef K. K., Nieto M. A (2024)** Epithelial–mesenchymal transition in tissue repair and degeneration. *Nat Rev Mol Cell Biol* **25**:720-739
25. **Mantovani A., Garlanda C. (2023)** Humoral Innate Immunity and Acute-Phase Proteins. *N Engl J Med* **388**:439-452
26. **Rojas A., Hachey W., Kaur G., Korejwo J., Muhammad R (2023)** Enhanced safety surveillance of STAMARIL® yellow fever vaccine provided under the expanded access investigational new drug program in the USA. *Journal of Travel Medicine* **30**:taad037
27. **Cottin P., Niedrig M., Domingo C (2013)** Safety profile of the yellow fever vaccine Stamaril®: a 17-year review. *Expert Review of Vaccines* **12**:1351-1368
28. **Chiu S. S., Lau Y.-L (2005)** Review of the VarilrixTM varicella vaccine. *Expert Review of Vaccines* **4**:629-643
29. **Van Den Ende C., Marano C., Van Ahee A., Bunge E. M., De Moerloose L (2017)** The immunogenicity and safety of GSK's recombinant hepatitis B vaccine in adults: a systematic review of 30 years of experience. *Expert Review of Vaccines* **16**:811-832
30. **Destexhe E., et al. (2013)** Evaluation of C-reactive protein as an inflammatory biomarker in rabbits for vaccine nonclinical safety studies. *Journal of Pharmacological and Toxicological Methods* **68**:367-373
31. **Tilbeurgh M. V., et al. (2021)** Predictive Markers of Immunogenicity and Efficacy for Human Vaccines. *Vaccines* **9**
32. **Vono M., et al. (2013)** The adjuvant MF59 induces ATP release from muscle that potentiates response to vaccination. *Proceedings of the National Academy of Sciences* **110**:21095-21100
33. **Calabro S., et al. (2011)** Vaccine adjuvants alum and MF59 induce rapid recruitment of neutrophils and monocytes that participate in antigen transport to draining lymph nodes. *Vaccine* **29**:1812-1823

34. Win Z., et al. (2021) Systematic Evaluation of Kinetics and Distribution of Muscle and Lymph Node Activation Measured by 18F-FDG- and 11C-PBR28-PET/CT Imaging, and Whole Blood and Muscle Transcriptomics After Immunization of Healthy Humans With Adjuvanted and Unadjuvanted Vaccines. *Front. Immunol* **11**
35. Mestas J., Hughes C. C. W (2004) Of Mice and Not Men: Differences between Mouse and Human Immunology. *J Immunol* **172**:2731-2738
36. Arunachalam A. B (2024) Vaccines Induce Homeostatic Immunity, Generating Several Secondary Benefits. *Vaccines* **12**
37. Torday J. S (2015) Homeostasis as the Mechanism of Evolution. *Biology* **4**:573-590
38. Koutsandreas T., Felden B., Chevet E., Chatziioannou A (2024) Protein homeostasis imprinting across evolution. *NAR Genomics and Bioinformatics* **6**:lqae014
39. Torzewski M., Waqar A. B., Fan J (2014) Animal Models of C-Reactive Protein. *Mediators of Inflammation* **683598**
40. Fu Y., Wu Y., Liu E (2020) C-reactive protein and cardiovascular disease: From animal studies to the clinic (Review). *Experimental and Therapeutic Medicine* **20**:1211
41. Fourati S., et al. (2022) Pan-vaccine analysis reveals innate immune endotypes predictive of antibody responses to vaccination. *Nat Immunol* **23**:1777-1787
42. Hagan T., et al. (2022) Transcriptional atlas of the human immune response to 13 vaccines reveals a common predictor of vaccine-induced antibody responses. *Nat Immunol* **23**:1788-1798
43. O'Connor D., et al. (2020) Gene expression profiling reveals insights into infant immunological and febrile responses to group B meningococcal vaccine. *Molecular Systems Biology* **16**
44. Reynaud K., Atwell S., Védrine C., Guyard C., Beloeil L (2025) Innovative micro physiological systems for vaccine development. *Human Vaccines & Immunotherapeutics* **21**
45. Gartlan K. H., et al. (2016) Sterile inflammation induced by Carbopol elicits robust adaptive immune responses in the absence of pathogen-associated molecular patterns. *Vaccine* **34**:2188-2196
46. Edgar R., Domrachev M., Lash A. E (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Research* **30**:207-210
47. Barrett T., et al. (2013) NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Research* **41**:D991-D995
48. Genome Reference Consortium (2010) Mus musculus genome assembly MGSCv37. NCBI. https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000001635.18/
49. Genome Reference Consortium (2009) Homo sapiens genome assembly GRCh37. NCBI. https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000001405.13/
50. (no date) Lift Genome Annotations. <https://genome.ucsc.edu/cgi-bin/hgLiftOver>
51. Benjamini Y., Hochberg Y. (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing - Benjamini. *Journal of the Royal Statistical Society: Series B (Methodological)* **57**:289-300
52. Meinshausen N., Bühlmann P. (2010) Stability selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **72**:417-473
53. Subramanian A., et al. (2005) Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences* **102**:15545-15550

Peer reviews

Reviewer #1 (Public review):

Summary:

The authors aimed to develop a translational framework for predicting vaccine reactogenicity by training a penalized ordinal regression model on mouse muscle transcriptomics and applying it across tissues and species to rank human vaccines by their inflammatory potential.

Strengths:

The study addresses an important gap in preclinical vaccine safety assessment. The identification of IL6/JAK/STAT3 signaling as a key pathway implicated in reactogenicity is biologically plausible, and the observation of coordinated changes between muscle and blood compartments supports the biological relevance of the signature. The model achieves near-perfect classification in mouse muscle tissue and successfully identifies Fluad (MF59-adjuvanted) as the most reactogenic among licensed human vaccines, consistent with clinical safety data.

Weaknesses:

The methodological foundation has several concerns. The reactogenicity class definitions rely on PC1 scores with modest variance explained, yet no sensitivity analyses demonstrate robustness to different normalization strategies, feature selection approaches, or dimensionality reduction methods. I suggest performing sensitivity analyses demonstrating that reactogenicity class definitions are robust to alternative normalization methods, feature selection criteria, and dimensionality reduction approaches.

The combined mouse analysis reveals that tissue effects dominate over vaccine-induced variation, and no explicit batch or compartment correction was reported. The authors can apply batch/compartment correction (e.g., SVA) when analyzing combined mouse muscle and blood data, then recompute PCA and downstream analyses.

The central claim regarding cross-species ranking capability is not fully supported. In human blood, the model largely distinguishes Fluad from other vaccines but shows limited separation among non-Fluad formulations, with many pairwise comparisons yielding non-significant adjusted p-values. This pattern suggests the model may be tuned to detect large inflammatory magnitudes-likely a consequence of training on extreme stimuli such as LPS and whole-cell pertussis-rather than capturing the finer gradations relevant for distinguishing licensed vaccines with moderate reactogenicity profiles. I highly suggest retraining the model, excluding extreme stimuli (LPS, Pentavac), to evaluate whether mid-range separations among licensed vaccines can be recovered.

Impact:

While the conceptual framework is promising, the current evidence does not convincingly demonstrate that the model can rank vaccines beyond identifying highly inflammatory outliers. The utility for preclinical assessment of novel vaccine candidates with moderate reactogenicity profiles remains uncertain.

<https://doi.org/10.7554/eLife.109928.1.sa1>

Reviewer #2 (Public review):

Summary:

The authors derived a time-specific signature of reactogenicity from mouse muscle following exposure to vaccines /TLRs for capturing the reactogenicity patterns. They tested this reactogenicity signature in mouse blood, and then they applied the reactogenicity signature to human blood from subjects having received different vaccines. They identified biomarkers

in mouse muscle which are also observed in mouse and human blood and could be used as a reactogenicity signature in mice, instead of CRP.

Strengths:

- (1) The authors used transcriptomic response following vaccination and used common genes to human and mice for defining a reactogenic signature.
- (2) As the authors used different formulations in mice, the model was trained across a broad reactogenicity spectrum, which has the advantage of being used for evaluating new vaccines/vaccine platforms.

Weaknesses:

- (1) The muscle gene signature reflects local reactogenicity. Systemic reactogenicity is not specifically addressed, except where overlapping gene signatures are observed for both local and systemic reactogenicity.
- (2) In the same logic, could we find additional genes in the blood which are not captured in the muscle?
- (3) The peak of the reactogenicity is usually 24h; it is not certain that additional TPs have helped the findings. If they have, the authors should explain.

<https://doi.org/10.7554/eLife.109928.1.sa0>