Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Uninfected BALB/c and C57BL/6 airways differ in *Tlr1/Tlr6* co-receptor and lipogenic transcripts, not adhesion-receptor pathways

Xiaotong Wang
Beijing Children's Hospital
Hao Chen
Beijing Children's Hospital
Chen Shen
shenchen1110@126.com

Beijing Children's Hospital

Research Article

Keywords: Mycoplasma pneumoniae, Mycoplasma pulmonis, inbred mouse strains, airway glycome, N- and O-glycans, Tlr1/Tlr6 co-receptor, lipogenesis

Posted Date: November 12th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11703601/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Abstract

BALB/c and C57BL/6 differ in their reported response to respiratory mycoplasmas, and the airway surface-receptor repertoire has been the usual explanation. We measured that repertoire directly in uninfected animals: released N- and O-glycans from lung and trachea of both strains (three per strain, 82 peaks), and the tracheal transcriptome re-quantified from raw alignments and analysed with DESeq2 (23). No candidate adhesion pathway could be resolved as different. Ciliary genes showed no enrichment among strain-divergent genes (12 of 782 against 12.7 expected after filtering polymorphic families, p = 0.49), and sulfatide synthesis, mucins and sialic acid enzymes were likewise unresolved. NeuAc residues per 100 glycans were higher in BALB/c in all four panels, while total sialylation was unchanged; the reciprocal fall in NeuGc holds in three panels but reverses in trachea O-glycans, so we report the NeuAc concordance and not a substitution, with whole-composition separation of both lung panels at the permutation floor available at this group size; no glycosyltransferase transcript accompanied it. Two differences outside the adhesion repertoire could be resolved. Transcript abundance at the Tlr1/Tlr6 co-receptor locus was higher in BALB/c (+ 2.20 and + 2.01), exceeding the envelope within which differential read mapping cannot be excluded and running opposite to its direction. A module of de novo lipogenesis transcripts was higher in BALB/c in all 13 members tested, at three hundred times that abundance, surviving controls for mapping bias, leukocyte fraction, adipose inclusion and adjacent-tissue carry-over. A second set of animals sequenced under a different condition reproduces 97% of evaluable gene-level directions and separates the two: the co-receptor transcript difference is unchanged there while the lipid module retains 0.39 of its magnitude.

Introduction

Inbred mouse strains differ in their reported response to respiratory mycoplasmas. For *Mycoplasma pulmonis*, strain-dependent differences in pulmonary burden and lesion severity are long established: mycoplasmal numbers in lung track genetic background rather than the state of adaptive immunity (1), and surfactant protein A-dependent, peroxynitrite-mediated killing by alveolar macrophages distinguishes resistant from susceptible animals (2). For experimental *Mycoplasma pneumoniae* infection, strain background likewise alters inflammatory and airway responses (3). Across this literature BALB/c is more often placed with the susceptible strains and C57BL/6 with the resistant; we take this as the motivating observation for the present study rather than as an established quantitative ranking, since the comparisons differ in organism, dose, route and readout, and we did not measure a susceptibility phenotype here.

The usual explanation for such differences has been the airway surface. Mycoplasmas are surface-adherent organisms, and *M. pneumoniae* binds sialylated poly-N-acetyllactosamine of the Ii-antigen series, a receptor class polarised in human bronchial epithelium to the apical and ciliary membranes of ciliated cells (4, 5); long-chain species are far more potent binding inhibitors than short-chain analogues (4). The sialic acid-binding site of the adhesion complex lies in P40/P90 rather than in P1 (6). On this

account a strain presenting more of the relevant receptor, or more ciliated surface on which to present it, should be the more readily colonised.

Strain-dependent glycan differences of the kind this account requires have been demonstrated in another organ and another pathogen. Intestinal mucin O-glycans differ in sialic acid content between inbred strains, the susceptible strain carrying less, and desialylation of mucins from the resistant strains increases their ability to block adherence of *Entamoeba histolytica* (7) — establishing both that inbred strains can differ in released O-glycan composition and that such differences can bear on pathogen attachment. For the mouse airway no equivalent comparison exists. Published mouse glycome resources profile N-glycans across organs of a single strain by MALDI-TOF after glycoblotting (8), and the intestinal comparison above used the same glycoblotting and MALDI-TOF route (7); both differ from the released-glycan fluorescence chromatography used here, and neither addresses the airway.

That account yields a prediction testable in uninfected animals: the susceptible strain should carry more of the relevant surface structures at steady state. It also carries a constraint. *M. pulmonis* does not use sialylated receptors; its colonies do not adsorb tracheal epithelial cells and neuraminidase treatment of host cells does not affect attachment (9). Its surface Vsa lipoproteins and EPS-I polysaccharide modulate adherence, but in a direction that does not map simply onto colonisation: mycoplasmas producing long Vsa proteins adhere poorly to epithelial cells yet colonise rodents proficiently (10). A single receptor-availability mechanism is therefore unlikely to cover both organisms, though we note this argument constrains a shared mechanism rather than excluding organism-specific ones.

We measured the airway glycan surface and the uninfected airway transcriptome of the two strains, asked whether the receptor prediction could be confirmed, and report what differed instead.

Results

No candidate adhesion pathway could be resolved as different

Among 782 genes annotated to ciliary biological processes, 12 reached a 5% false-discovery rate against 19.1 expected by chance — no enrichment. Applying the same polymorphic-family filter used throughout this study (333 of 535 significant genes remain) the expectation falls to 12.7 against the same 12 observed (hypergeometric p = 0.49), so there is no evidence of depletion either. *Foxj1*, which specifies ciliated cell fate, was indistinguishable between strains (log2 fold change + 0.04; 95% CI 0.62- to 1.71-fold; baseMean 1034), as were axonemal dyneins, radial spoke components and basal body genes.

Sulfatide, the best-characterised non-sialylated receptor for *M. pneumoniae*, is a glycolipid and outside what enzymatic release of protein-linked glycans detects; we tested its biosynthetic pathway instead. Neither *Gal3st1* nor *Ugt8a*, *Ugcg*, *Arsa*, *Sulf1* or *Sulf2* reached significance. The sialic acid-modifying

enzymes *Cmah*, *Neu1*, *Neu3*, *St3gal4* and *St6gal1* were likewise unresolved. Among mucins, *Muc5b* — the dominant airway mucin at baseMean 156,094 — was 0.56-fold in BALB/c (95% CI 0.25 to 1.23) and *Tff2* 0.28-fold; neither reaches significance at this sample size, but both point downward and neither should be described as equivalent between strains.

Throughout this section the correct reading is that three animals per group could not resolve these pathways, not that they are equivalent. Genome-wide, 1,126 genes have |log2 fold change| > 1 and remain non-significant in this dataset, which bounds what a non-significant result here is worth.

In lung O-glycans the strain difference is a sialic acid species substitution

Of 82 chromatographic peaks, 66 carried evaluable data in both strains. One structure differed under all three statistical treatments applied: the disialylated core-1 O-glycan HexHexNAc(NeuAc)2, higher in BALB/c lung (16.49% versus 12.60%). Decomposed by sialic acid species, NeuAc-disialylated species were higher in BALB/c (+ 3.35) and NeuGc-disialylated species lower (− 1.80), while total sialylation was similar (77.37% versus 78.78%): a substitution rather than an increase. The three treatments are not independent evidence; their p-values correlate at Spearman 0.99 (Student versus Welch) and 0.87 (Student versus compositional log-ratio) — so this is a consistency check, not triangulation. Because the exact permutation floor for three animals per group is p = 0.10 two-sided, these are reported as effect sizes and the accompanying q values are not treated as inferential.

No O-glycan in either tissue extended beyond (Hex)2(HexNAc)2(NeuAc)2. The long poly-N-acetyllactosamine backbone that constitutes the *M. pneumoniae* receptor was therefore absent from the dataset, and *B3gnt2*, the elongase that builds it, was flat (log2 fold change − 0.15, adjusted p = 1.00). We can state that this receptor class was not detected; we cannot state that it differs between strains.

Two patterns in the glycan data are not captured by the per-peak screen. First, the whole composition of both lung panels separates the strains: an exact permutation test on Aitchison distances over all ten distinct three-against-three splits ranks the true split first in lung N-glycans (pseudo-F 7.35 against 1.04 for the next-best split) and first in lung O-glycans (15.13 against 1.11), each reaching p = 0.10, the floor available at this group size. Neither tracheal panel separates. Second, NeuAc residues per 100 glycans are higher in BALB/c in all four panels (lung N + 2.58, trachea N + 2.56, lung O + 5.68, trachea O + 1.00; binomial P for four of four = 0.125 two-sided). The reciprocal fall in NeuGc that a substitution would require holds in three panels (lung N -1.39, trachea N -4.97, lung O -5.06) but reverses in trachea O-glycans (+ 2.83), so we describe a concordant NeuAc excess and not a substitution. The joint permutation statistic does not rank the true split first, and we report the concordance rather than a p value for it. Neither pattern has a transcript correlate: *Cmah*, the hydroxylase whose loss would produce the lung O-glycan substitution, is unresolved (log2 fold change − 0.10, adjusted p = 1.00, 95% confidence interval 0.48- to 1.81-fold), as is the CMP-sialic acid transporter *Slc35a1* (-0.02). The glycan and transcript layers remain uncoupled.

Differences reproduce in a second set of animals under a different condition

The submitted expression matrix contains six further animals of the two strains that were sequenced under a different condition — three BALB/c and three C57BL/6 taken three days after *M. pneumoniae* inoculation — assignable by strain from the *Nnt* deletion that distinguishes C57BL/6J (21) (11) (FPKM 8.6–12.6 in BALB/c against 4.4–5.9 in the C57BL/6 background). Because our own analysis of these animals resolved no main effect of infection at any gene, the strain contrast within them serves as a within-study check rather than as a replication under identical conditions, and we describe it as such. Of the 333 filtered significant genes, 261 yield an evaluable ratio in both conditions, and 252 of these keep their sign (96.6%, binomial p = 3.8e-63), with effect sizes correlated across genes (Pearson r = 0.840, p = 1.3e-70, n = 261). The two conditions nonetheless separate the findings reported below. Because the second set is available only as FPKM, this comparison uses FPKM group-mean ratios throughout rather than the shrunken DESeq2 estimates quoted elsewhere; on that scale the baseline co-receptor transcript difference is + 1.44 for *Tlr1* and + 0.93 for *Tlr6* (DESeq2: +2.20 and + 2.01). At the co-receptor locus the effect is unchanged between conditions (+ 1.56 and + 0.81 in the second set, median retention 0.98), whereas the lipid module described next retains only 0.39 of its baseline magnitude on the same scale.

Transcript abundance at the Tlr1/Tlr6 locus is higher in BALB/c

Re-quantification of the uninfected tracheal transcriptome from raw alignments identified 535 genes at a 5% false-discovery rate among 21,897 tested. Applying an explicit symbol filter for strain-polymorphic and copy-number-variable families (predicted genes, pseudogenes, ribosomal-protein pseudogenes, the H2 haplotype block, Klra, Skint and Raet1 clusters, and *Nnt*, *Erdr1*, *Eno1b*; regular expression given in Methods) removes 202, leaving 333.

Within that set, *Tlr1* and *Tlr6* were both higher in BALB/c (+ 2.20 and + 2.01; adjusted p = 2.3×10^{-6} and 1.0×10^{-3}), as was *Lbp* (+ 1.42, adjusted p = 5.9×10^{-5}). *Tlr1* and *Tlr6* lie 12 kb apart and constitute one locus, not two observations. *Tlr2*, the obligate signalling subunit through which both accessory chains act, was not resolved (+ 0.50; 95% CI 0.85- to 2.34-fold, which admits a two-fold increase), nor were *Cd14*, *Myd88*, *Irak1*, *Irak4*, *Traf6*, *Tirap*, *Nfkb1* or *Rela*. The difference is confined to the accessory subunits; whether the pathway they feed also differs is beyond what this sample size can determine. *Tlr6* is expressed at low abundance (baseMean 62 against 115 for *Tlr1* and 1,609 for *Lbp*), so its estimate is the least precise of the three.

This result survives the two artefacts that dominate a two-strain comparison, and does so because both act in the opposite direction. First, reads were aligned to a C57BL/6 reference, so sequence divergence biases BALB/c downward: among seven loci of known pure divergence, six are negative (*H2-K1* − 1.69, *H2-Q7* − 1.21, *H2-Eb1* − 1.18, *H2-Ab1* − 0.93, *H2-D1* − 0.64, *B2m* − 0.20) and only *Nnt* is positive (+ 1.11, reflecting the C57BL/6 deletion). Both co-receptor genes, and *Lbp*, exceed that envelope and run upward. Second, a 24-gene myeloid and leukocyte reference panel was lower in BALB/c (median log2 fold

change − 0.25, Wilcoxon p = 0.0065), indicating a roughly 16% lower leukocyte fraction; this too would depress, not elevate, these transcripts. Neither artefact can generate an upward difference of this size.

The direction of the co-receptor transcript difference is reproduced in a purified cell population. In published transcriptomes of unstimulated tissue-resident macrophages from the same two strains (GSE248033, three animals per strain), *Tlr1* and *Tlr6* were again higher in BALB/c (+ 0.32, p = 0.007; +0.68, p = 0.001). That dataset is a macrophage-enriched peritoneal cell population rather than a pure one: macrophage and myeloid transcripts dominate (*Lyz2* 15.5, *Itgam* 12.5, *Adgre1* 11.5 log2 CPM) but B-cell transcripts are also present (*Ms4a1* 8.6, *Cd19* 7.2), while airway epithelium is absent (*Epcam* 2.5; *Krt5*, *Scgb1a1* and *Foxj1* undetected). Because it contains no airway epithelium and no submucosal gland, the leukocyte-fraction difference present in our whole-tissue data cannot generate the co-receptor transcript difference there. We describe this as a cell-intrinsic consistency check rather than independent validation, for two reasons. That dataset was also aligned to a C57BL/6 reference and so shares the mapping bias, and the agreement does not extend beyond this pair: of nine genes measurable in both datasets only five agree in sign, which is what chance predicts (binomial p = 0.50), and effect sizes are uncorrelated across the nine (Pearson r = 0.06, p = 0.87). *Lbp* in particular reverses (− 2.60, p = 0.0002), which we attribute to cell source — *Lbp* is secreted predominantly by liver and epithelium — and which is why we do not carry *Lbp* forward as part of the result.

A lipid-synthesis module is higher in BALB/c, but is condition-dependent

The largest coherent signal among the filtered genes is metabolic rather than immune. A two-sided signed-rank test on the thirteen members returns p = 0.00024, but this is the exact floor of the test at n = 13 (2/2^13) and is the same for any thirteen same-sign values, so it carries no information about effect size; treating the module as five independent units gives a sign-test p of 0.062544): *Pdk4* + 1.77, *Thrsp* + 1.68, *Elovl6* + 1.48, *Hmgcs2* + 1.48, *Angptl4* + 1.47, *Acly* + 1.44, *Pygl* + 1.39, *Acacb* + 1.35, *Gpd1* + 1.32 and *Fabp4* + 1.27. These are abundant transcripts: the significant members sum to a baseMean of 53,619, against 177 for *Tlr1* and *Tlr6* together.

Four controls separate this from the artefacts that defeated other candidate modules. All ten significant members exceed the upward limit of the sequence-divergence envelope (+ 1.11), the direction opposite to the one mapping bias can produce. The myeloid reference panel runs downward (median − 0.36) while this module runs upward. A twelve-gene adipocyte panel (*Adipoq*, *Cfd*, *Plin1*, *Cidec*, *Lep*, *Retn*, *Pparg*, *Car3*, *Adrb3*, *Lipe*, *Slc7a10*, *Fabp4*) is not uniformly flat: nine of twelve run positive with a median of + 0.35, though only *Fabp4* reaches significance. Excluding *Fabp4*, which is also one of the ten significant module members and so cannot serve as both control and result, the remaining eleven give a median of + 0.34 and a signed-rank p of 0.24. Adipose contamination is therefore not excluded outright by this panel; what argues against it is that the canonical adipocyte markers *Adipoq* (-0.05), *Cidec* (+ 0.25) and *Plin1* (+ 0.69) stay far below the module members (median + 1.46) despite comparable abundance. And none of the

lipid transcripts covaries with the thyroid or skeletal-muscle markers that index adjacent-tissue carry-over (all |r| <= 0.60, all p > = 0.21). Circadian transcripts are uniformly unresolved (14 genes, all padj = 1.00), excluding a difference in sampling time.

Two features qualify the result. Fatty-acid oxidation markers move in the same direction as the synthetic enzymes (*Pdk4*, *Hmgcs2*, *Angptl4*, *Acadm* all up), which is not the reciprocal relationship expected if a single anabolic programme were engaged, and no master regulator is resolved (*Srebf1* 0.00, *Mlxipl* + 0.98 at p = 0.060, *Ppara* + 0.39, all padj > = 0.06). More importantly, the module is not constitutive in the way the co-receptor locus is: on the FPKM scale used for that comparison it retains only 0.39 of its baseline magnitude in the second set of animals, so it describes a metabolic state that infection modifies rather than a fixed property of the strains. We report it as the largest difference in the dataset while noting that it is the less stable of the two.

Antimicrobial transcripts are lower in BALB/c, but not separably from composition

Seven antimicrobial transcripts were lower in BALB/c, spanning six independent loci after collapsing the chr6 *Reg* cluster. We report this as an observation rather than a finding, because unlike the co-receptor result it is not separable from the two artefacts. All four of the abundant members (*Lgals3* − 1.36, *Pycard* − 1.31, *Reg3g* − 1.23, *Ccl6* − 1.02) fall inside the sequence-divergence envelope and run in the same direction as the mapping bias, and the 16% lower leukocyte fraction accounts for part of the shift. *Reg3g*, the most abundant of them, is predominantly an intestinal and gastric transcript and sits within a block of gastro-oesophageal carry-over in these dissections. Protein-level measurement would be required to establish this as a regulatory difference.

Discussion

What a transcript difference at this locus can support is narrow. TLR1 and TLR6 reach the cell surface only in complex with TLR2, and ligand discrimination depends on that assembly rather than on the abundance of either chain alone. *Tlr2* itself is unresolved here (95% CI 0.86- to 2.30-fold), and no protein level, surface complex, dimerisation state or ligand response was measured. The observation is therefore that two accessory chains carry a transcript difference while the shared transduction core does not; whether that difference reaches the receptor complex is untested.

The receptor account of strain-dependent mycoplasma response makes a prediction testable without infecting an animal, and we could not confirm it in any compartment we measured. Cilia, sulfatide synthesis, mucins and sialic acid-modifying enzymes were all unresolved between strains, and the receptor class *M. pneumoniae* binds was not detected in the glycan dataset at all. The honest statement is that three animals per group could not resolve these pathways. We note one feature that nonetheless weighs against a simple receptor explanation: the organism-level constraint that *M. pulmonis* produces comparable strain rankings without using sialylated receptors means a shared receptor-availability

mechanism would have to operate through two unrelated binding systems. We do not claim that nothing was elevated in the susceptible strain. Four peaks were higher in BALB/c under the equal-variance test, including a sialylated fucosylated lung N-glycan (H6N4SaF, + 1.07 and + 1.40 at two peaks), which belongs to a candidate receptor class; none survived all three treatments, and we report them here rather than leave the impression of a uniformly downward picture. The interest of this locus is that both mycoplasmas lack a cell wall and present diacylated lipoproteins as their dominant pattern, and that the *M. pulmonis* Vsa proteins are themselves lipoproteins, so the molecular class these accessory chains discriminate is one the two organisms share. TLR2 is the obligate signalling subunit for both diacylated and triacylated lipopeptide; TLR6 and TLR1 are the accessory chains that determine which acylation class is recognised (12). A difference confined to the accessory pair is therefore the shape a ligand-class-specific difference would take — but transcript abundance is not sensing capacity, *Tlr2* itself was not resolved here, and no surface protein, receptor-complex or signalling readout was measured.

We also emphasise what this result does not establish. Published work shows that reconstituting airway epithelial TLR2 improves *M. pneumoniae* clearance (13), which means more signalling through this pathway is protective. Our observation that the strain usually described as susceptible carries higher accessory-chain transcripts therefore does not fit a simple more-signalling-is-worse model, and we do not propose one. Possible reconciliations — that accessory-chain abundance does not track complex availability when the obligate subunit is unchanged, or that the relationship between sensing and outcome is non-monotonic — are testable but not tested here. Independent work also reports that neutralising IL-17A reduces *M. pulmonis* disease severity in BALB/c without changing bacterial burden (14), consistent with response rather than burden governing severity in this system, though that study does not address the present locus.

One cross-pathogen comparison is worth stating, together with why it does not test the present locus. Where mouse strain susceptibility to a mycobacterium has been measured directly, BALB/c and C57BL/6 fall in the same resistant group (15), and mycobacterial lipoproteins are TLR2/TLR1 agonists, so the absence of a strain difference there might seem to argue against the locus mattering. It does not, because outcome in murine tuberculosis is not set at this step: mice lacking TLR2, TLR4 and TLR9 together contain aerosol *M. tuberculosis* infection, expressing pro-inflammatory cytokines, antigen-specific IFN-γ and IFN-γ-dependent iNOS comparably to wild type (16). Control of a slow-growing intracellular mycobacterium is rate-limited by Th1 polarisation and phagosomal killing rather than by lipopeptide recognition, so a co-receptor difference would not be expected to change it. We note this because it illustrates the general point: whether transcript abundance at a sensing locus translates into an outcome difference depends on whether sensing is rate-limiting for the organism in question, and we have not shown that it is for either mycoplasma.

Several limitations bound this study. Three animals per group leave an exact permutation floor of p = 0.10 two-sided for the glycan comparisons; the DESeq2 analyses are parametric and borrow dispersion across genes, so that floor does not apply to them, but their intervals remain wide and we report intervals for every null claim. The dissections contain substantial adjacent tissue: thyroid transcripts

account for 8.3% of counts (*Tg* baseMean 1,624,371) and skeletal muscle for 2.9%, against *Scgb1a1* at 6,515 for the airway itself. This carry-over is randomly distributed with respect to strain — *Tg* ranges 2,899 to 7,506 across both groups — and none of the genes reported here covaries with it (all |Spearman| ≤ 0.54, p ≥ 0.27), so it reduces sensitivity rather than creating strain differences; but it bounds how finely airway-specific signals can be read. The transcriptome is tracheal while the significant glycan difference is in lung. The glycan method sees protein-linked glycans only, so glycolipid receptors are invisible to it, and reports composition rather than linkage, so the α2,3 linkage relevant to *M. pneumoniae* cannot be assessed. We measured only uninfected animals and did not measure a susceptibility phenotype, so nothing here tracks susceptibility.

Several pathways a reader would expect to differ between these strains did not. No gene of the Th1/Th2 polarisation axis, the IL-17 axis, the nitric oxide and peroxynitrite pathway, or the surfactant proteins *Sftpa1* and *Sftpd* reached significance, and neither did *Slc11a1*, the classical inbred-strain resistance locus. At three animals per group these are failures to resolve rather than demonstrations of equivalence, and we list them so that their absence from the Results is not mistaken for their not having been examined.

A further class of apparent differences was excluded rather than reported. Transcripts of submucosal and salivary gland origin (*Lpo*, *Sbpl*, *Dcpp3*, *Azgp1*, *Smr2*) showed very large downward shifts at high abundance, but pairwise correlations among them are 0.94–0.97 and a single latent factor accounts for 97% of their joint variance, a factor that predicts each gene within strains as well as between them: they index how much gland was included in the dissection, not a property of the strains. Several large categories reaching nominal significance — neuronal and synaptic transcripts, vesicle traffic, ribosomal genes, adhesion and matrix — run downward and lie within or below the sequence-divergence envelope at low abundance, the configuration mapping bias produces, and are likewise not reported as findings.

The locus is testable. The informative experiments are to measure TLR1/TLR6 protein and surface complex in the two strains, to compare lipopeptide responsiveness of tracheal explants or airway cells, and to infect *Tlr6*-deficient animals on each background.

Materials and Methods

Animals. Male BALB/c and C57BL/6J mice were housed under specific-pathogen-free conditions. The primary comparison used three uninfected animals per strain. A further three animals per strain, from the same colony and sequenced in the same batch, had received *M. pneumoniae* by the route and dose described in the originating study and were taken three days after inoculation; these contribute only to the within-study check reported in the Results and to no other analysis. Trachea and lung were collected at the same session. Tracheal dissections were whole-tissue and included adjacent structures: by marker transcript abundance the sequenced material comprises approximately 8.3% thyroid (*Tg* baseMean 1,624,371), 2.9% skeletal muscle (*Ttn* 436,051) and 0.7% oesophageal or gastric tissue (*Lipf* 147,340), with airway-specific markers together accounting for roughly 3.1%. Within the airway

component, submucosal gland and mucous transcripts dominate (*Ltf* 146,654; *Muc5b* 156,094) over secretory epithelium (*Bpifa1* 50,667) and ciliated epithelium (*Dnah5* 3,713). All animal procedures, including the infection of the second set of animals, were approved by the Animal Ethics Center of Beijing Laboratory Animal Research Center (protocol **BLARC-LAWER-202307002)** and performed in accordance with institutional and national guidelines.

Glycan profiling. N- and O-glycans were released from lung and trachea of each animal, fluorescently labelled, and separated by liquid chromatography with on-line fluorescence and mass-spectrometric detection (Waters Xevo G2-XS quadrupole time-of-flight mass spectrometer with an ACQUITY fluorescence detector, MassLynx acquisition). Twenty-eight acquisitions were performed, one per tissue-by-glycan-class panel per animal, over three consecutive days. Quantification is from the fluorescence channel (ACQUITY FLR, excitation 310 nm, emission 370 nm, the standard setting for 2-aminobenzamide labelling (17); the labelling reagent is not recorded in the acquisition metadata available to us); relative abundance is fluorescence peak area as a percentage of the panel total. N-glycan separations ran for 50 min and O-glycan separations for 32 min, both sampled at 0.1 s. Mass spectrometry was acquired in positive-ion electrospray mode with a 0.5 s scan time, centroided, with sodium formate calibration, using a single TOF-MS function plus a lock-mass reference function and a transfer collision energy of 6 eV; the method therefore contains no fragmentation channel, and structures are assigned by composition and retention time rather than by tandem mass spectrometry.

Peak assignments were checked against the raw chromatograms: of the twelve lung O-glycan peaks, nine fell within 0.2 min of the reported retention time (median deviation 0.091 min), and independent integration of the fluorescence trace correlated with the reported areas at Pearson r = 0.84. Three deviations were identified and are carried into the Results and limitations. The vendor report flags peaks 3 and 4 (HexNeuAc, HexNeuGc) as possible sample-preparation by-products; these also lack HexNAc and so are not mucin-type O-glycans, and a sensitivity analysis excluding them is reported. The two entries sharing the composition HexHexNAcNeuAcNeuGc are not baseline-resolved but a single peak with a shoulder, so only their sum is used. One tracheal N-glycan acquisition terminated at 37.96 min instead of 50 min, which accounts for the higher proportion of missing values in that panel.

Peaks lacking evaluable data in either strain were excluded, leaving 66. Group comparisons used Student and Welch t-tests and a compositional centred log-ratio transform (18), with Benjamini-Hochberg correction (19) within panel. Panel-level composition was compared by permutational multivariate analysis on Aitchison distances (20), enumerating all ten distinct three-against-three splits exactly.

Transcriptome. Tracheal RNA-seq alignments to GRCm38 were re-quantified with featureCounts (21) against GENCODE (22) M25 (exon level, gene_name attribute). Genes with at least 10 counts in three or more samples were retained (22,043 of 55,291; 21,897 returned a test statistic) and tested with DESeq2 (23) using a strain term, BALB/c relative to C57BL/6. Strain-polymorphic and copy-number-variable families were flagged by the regular expression ^Gm\d+$|Rik$|-ps[0–9]{0,3}$|-rs\d+$|^Rp[sl]\d.*\d$|^H2-|^Klra|^Skint|^Raet1|^Nnt$|^Erdr1$|^Eno1b$|^Trim(?:12|30|34) applied to gene symbols, which removes

202 of the 535 significant genes. The expression is given verbatim; the asterisk quantifier in ^Rp[sl]\d.*\d$ is load-bearing and is lost if the text is reflowed by a word processor. The flagged column deposited in the source data reflects the earlier form of this expression and marks 178; the 24 additional Riken clones removed here are recoverable from the gene symbols.

Artefact controls. Two artefacts dominate a two-strain comparison aligned to a single reference. Differential mapping was assessed by benchmarking against seven loci of known pure sequence divergence between these strains (*H2-K1*, *H2-Q7*, *H2-Eb1*, *H2-Ab1*, *H2-D1*, *B2m*, *Nnt*), whose effect sizes define an envelope of − 1.69 to + 1.11 with six of seven negative, consistent with reads from the non-reference strain being lost. Cell-composition difference was assessed with a 24-gene myeloid and leukocyte reference panel compared by Wilcoxon signed-rank test against zero. Extraneous-tissue carry-over was quantified from marker transcripts and its association with reported genes tested by Spearman correlation across the six animals.

Gene set analysis. Biological process annotations were taken from the mouse GO annotation file (24). Gene sets of 10 to 600 members present in the tested background were evaluated by hypergeometric test with Benjamini-Hochberg correction. Ciliary gene sets were identified by membership of core ciliary genes rather than by term name.

Statistics. With three animals per group, an exact permutation test over the 20 group assignments cannot return a two-sided p below 0.10; this bound applies to the glycan comparisons, which are per-peak two-group tests on six exchangeable units, and those q values are presented as descriptive orderings. It does not apply to the DESeq2 analyses, which use a parametric Wald test with dispersion shared across genes; for those we report 95% confidence intervals for the genes named individually as unresolved; for gene panels declared unresolved we report that no member reached significance and provide the full table rather than an interval for each. Linked genes were collapsed to loci before any statement about the number of independent observations.

Declarations

Competing interests.

The authors declare no competing interests.

Use of AI.

AI tools were used for language editing and for assistance with code generation and modification; the author takes full responsibility for the content.

Funding.

This work was supported by the Beijing Science and Technology Nova Program Interdisciplinary Project (20230484439) and the Beijing High-Level Public Health Technical Talents Project (2023-02-33). The funder had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Author contributions.

Xiaotong Wang[1†] and Hao Chen[1†] (co-first authors): animal husbandry and breeding, specimen collection, testing, data analysis, writing — original draft. Chen Shen[1*]: conceptualization, study design, writing — review and editing, funding acquisition, supervision, and accountability for the integrity of the research. Writing — review and editing: Xiaotong Wang, Hao Chen and Chen Shen.

Data availability.

All processed data supporting the conclusions are provided as source data with this article: per-peak relative abundances for all 82 chromatographic peaks, the per-animal abundance matrix, the DESeq2 result table for all 21,897 tested genes with the polymorphic-family flag (the deposited file must contain every tested gene, not only the 535 significant ones, so that each unresolved claim can be checked), gene set enrichment results, and the artefact-control panels.

References

1. Cartner SC, Lindsey JR, Gibbs-Erwin J, Cassell GH, Simecka JW (1998) Roles of innate and adaptive immunity in respiratory mycoplasmosis. Infect Immun 66:3485–3491. https://doi.org/10.1128/IAI.66.8.3485-3491.1998
2. Hickman-Davis J, Gibbs-Erwin J, Lindsey JR, Matalon S (1999) Surfactant protein A mediates mycoplasmacidal activity of alveolar macrophages by production of peroxynitrite. Proc Natl Acad Sci U S A 96:4953–4958. https://doi.org/10.1073/pnas.96.9.4953
3. Fonseca-Aten M, Ríos AM, Mejías A, Chávez-Bueno S, Katz K, Gómez AM, McCracken GH, Hardy RD (2005) Mycoplasma pneumoniae induces host-dependent pulmonary inflammation and airway obstruction in mice. Am J Respir Cell Mol Biol 32:201–210. https://doi.org/10.1165/rcmb.2004-0197OC
4. Loomes LM, Uemura K, Feizi T (1985) Interaction of Mycoplasma pneumoniae with erythrocyte glycolipids of I and i antigen types. Infect Immun 47:15–20. https://doi.org/10.1128/iai.47.1.15-20.1985
5. Loveless RW, Feizi T (1989) Sialo-oligosaccharide receptors for Mycoplasma pneumoniae and related oligosaccharides of poly-N-acetyllactosamine series are polarized at the cilia and apical-microvillar domains of the ciliated cells in human bronchial epithelium. Infect Immun 57:1285–1289. https://doi.org/10.1128/iai.57.4.1285-1289.1989

6. Vizarraga D, Kawamoto A, Matsumoto U, Illanes R, Pérez-Luque R, Martín J, Mazzolini R, Bierge P, Pich OQ, Espasa M, Sanfeliu I, Esperalba J, Fernández-Huerta M, Scheffer MP, Pinyol J, Frangakis AS, Lluch-Senar M, Mori S, Shibayama K, Kenri T, Kawakita Y, Miyata M, Serrano L, Fita I (2020) Immunodominant proteins P1 and P40/P90 from human pathogen Mycoplasma pneumoniae. Nat Commun 11:5188. https://doi.org/10.1038/s41467-020-18777-y
7. Kato K, Takegawa Y, Ralston KS, Gilchrist CA, Hamano S, Petri WA, Shinohara Y (2013) Sialic acid-dependent attachment of mucins from three mouse strains to Entamoeba histolytica. Biochem Biophys Res Commun 436:252–258. https://doi.org/10.1016/j.bbrc.2013.05.085
8. Otaki M, Hirane N, Natsume-Kitatani Y, Nogami Itoh M, Shindo M, Kurebayashi Y, Nishimura SI (2022) Mouse tissue glycome atlas 2022 highlights inter-organ variation in major N-glycan profiles. Sci Rep 12:17804. https://doi.org/10.1038/s41598-022-21758-4
9. Sobeslavsky O, Prescott B, Chanock RM (1968) Adsorption of Mycoplasma pneumoniae to neuraminic acid receptors of various cells and possible role in virulence. J Bacteriol 96:695–705. https://doi.org/10.1128/jb.96.3.695-705.1968
10. Bolland JR, Dybvig K (2012) Mycoplasma pulmonis Vsa proteins and polysaccharide modulate adherence to pulmonary epithelial cells. FEMS Microbiol Lett 331:25–30. https://doi.org/10.1111/j.1574-6968.2012.02551.x
11. Freeman HC, Hugill A, Dear NT, Ashcroft FM, Cox RD (2006) Deletion of nicotinamide nucleotide transhydrogenase: a new quantitative trait locus accounting for glucose intolerance in C57BL/6J mice. Diabetes 55:2153–2156. https://doi.org/10.2337/db06-0358
12. Takeda K, Takeuchi O, Akira S (2002) Recognition of lipopeptides by Toll-like receptors. J Endotoxin Res 8:459–463. https://doi.org/10.1179/096805102125001073
13. Wu Q, Jiang D, Minor MN, Martin RJ, Chu HW (2011) In vivo function of airway epithelial TLR2 in host defense against bacterial infection. Am J Physiol Lung Cell Mol Physiol 300:L579–L586. https://doi.org/10.1152/ajplung.00336.2010
14. Mize MT, Sun XL, Simecka JW (2018) Interleukin-17A exacerbates disease severity in BALB/c mice susceptible to lung infection with Mycoplasma pulmonis. Infect Immun 86:e00292–e00218. https://doi.org/10.1128/IAI.00292-18
15. Medina E, North RJ (1998) Resistance ranking of some common inbred mouse strains to Mycobacterium tuberculosis and relationship to major histocompatibility complex haplotype and Nramp1 genotype. Immunology 93:270–274. https://doi.org/10.1046/j.1365-2567.1998.00419.x
16. Hölscher C, Reiling N, Schaible UE, Hölscher A, Bathmann C, Korbel D, Lenz I, Sonntag T, Kröger S, Akira S, Mossmann H, Kirschning CJ, Wagner H, Freudenberg M, Ehlers S (2008) Containment of aerogenic Mycobacterium tuberculosis infection in mice does not require MyD88 adaptor function for TLR2, -4 and – 9. Eur J Immunol 38:680–694. https://doi.org/10.1002/eji.200736458
17. Royle L, Campbell MP, Radcliffe CM, White DM, Harvey DJ, Abrahams JL, Kim YG, Henry GW, Shadick NA, Weinblatt ME, Lee DM, Rudd PM, Dwek RA (2008) HPLC-based analysis of serum N-glycans on a

96-well plate platform with dedicated database software. Anal Biochem 376:1–12. https://doi.org/10.1016/j.ab.2007.12.012

18. Gloor GB, Wu JR, Pawlowsky-Glahn V, Egozcue JJ (2016) It's all relative: analyzing microbiome data as compositions. Ann Epidemiol 26:322–329. https://doi.org/10.1016/j.annepidem.2016.03.003
19. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B Stat Methodol 57:289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x
20. Anderson MJ (2001) A new method for non-parametric multivariate analysis of variance. Austral Ecol 26:32–46. https://doi.org/10.1111/j.1442-9993.2001.01070.pp.x
21. Liao Y, Smyth GK, Shi W (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30:923–930. https://doi.org/10.1093/bioinformatics/btt656
22. Frankish A, Carbonell-Sala S, Diekhans M, Jungreis I, Loveland JE, Mudge JM, Sisu C, Wright JC, Arnan C, Barnes I, Banerjee A, Bennett R, Berry A, Bignell A, Boix C, Calvet F, Cerdán-Vélez D, Cunningham F, Davidson C, Donaldson S, Dursun C, Fatima R, Giorgetti S, Giron CG, Gonzalez JM, Hardy M, Harrison PW, Hourlier T, Hollis Z, Hunt T, James B, Jiang Y, Johnson R, Kay M, Lagarde J, Martin FJ, Gómez LM, Nair S, Ni P, Pozo F, Ramalingam V, Ruffier M, Schmitt BM, Schreiber JM, Steed E, Suner MM, Sumathipala D, Sycheva I, Uszczynska-Ratajczak B, Wass E, Yang YT, Yates A, Zafrulla Z, Choudhary JS, Gerstein M, Guigo R, Hubbard TJP, Kellis M, Kundaje A, Paten B, Tress ML, Flicek P (2023) GENCODE: reference annotation for the human and mouse genomes in 2023. Nucleic Acids Res 51:D942–D949. https://doi.org/10.1093/nar/gkac1071
23. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550. https://doi.org/10.1186/s13059-014-0550-8
24. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G (2000) Gene ontology: tool for the unification of biology. Nat Genet 25:25–29. https://doi.org/10.1038/75556

Figures

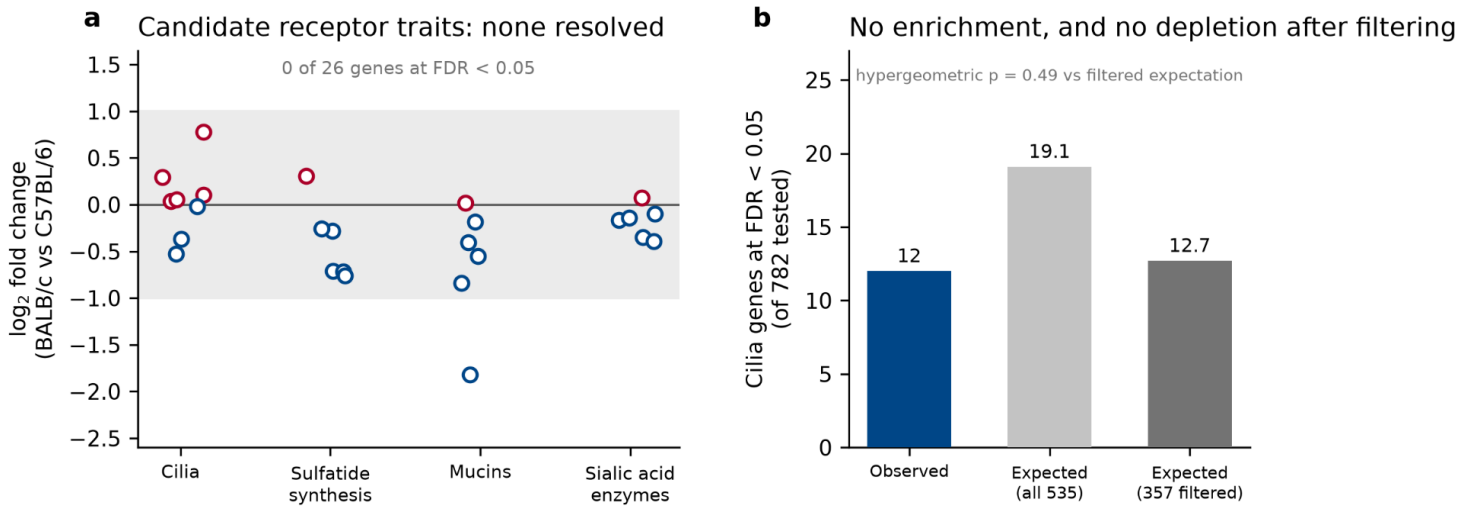


Figure 1

Candidate adhesion pathways are not resolved as different between the strains.

(a) Log2 fold change for 26 transcripts of four candidate receptor and adhesion categories — ciliary genes, sulfatide synthesis, mucins and sialic acid enzymes — BALB/c against C57BL/6. None reached a 5% false-discovery rate. The shaded band marks |log2 fold change| < 1. (b) Ciliary genes reaching significance, observed against the expectation under the unfiltered gene set and under the polymorphic-family filter applied throughout this study; the filtered comparison gives hypergeometric p = 0.49.

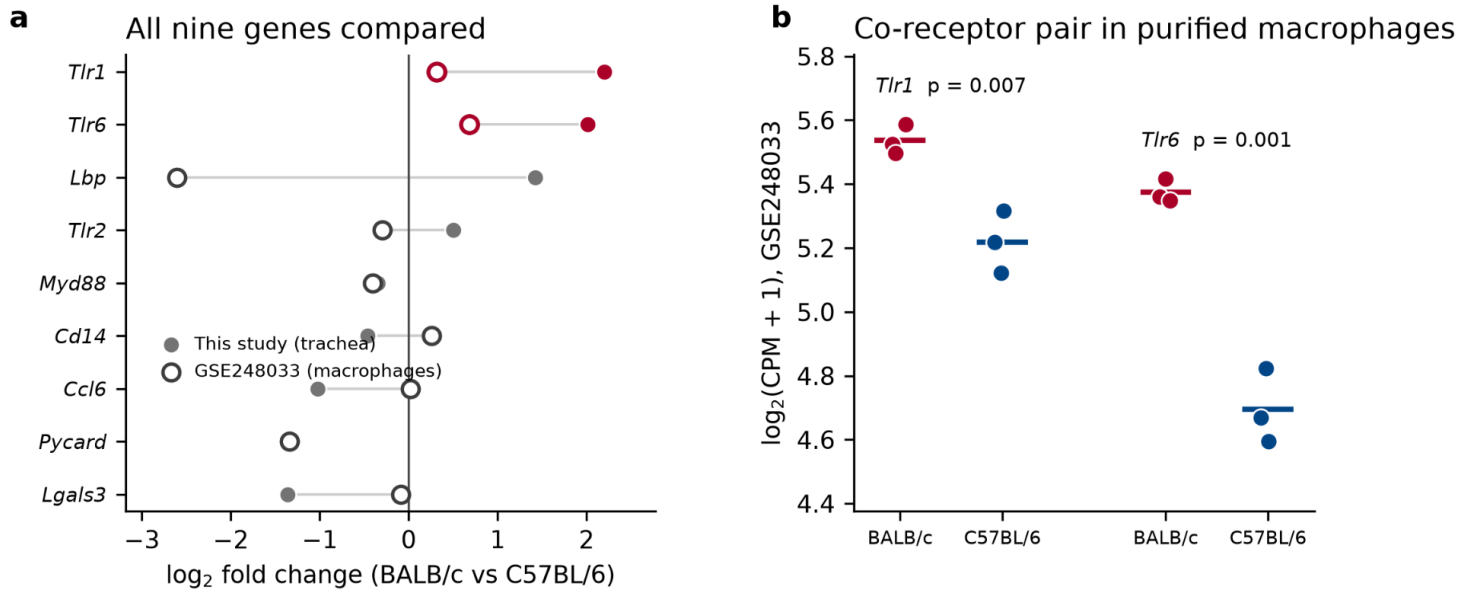


Figure 2

The co-receptor transcript difference in an independent macrophage-enriched dataset.

(a) All nine genes measurable in both datasets, this study against GSE248033. Sign agreement is five of nine, which chance predicts, and effect sizes are uncorrelated; Tlr1 and Tlr6 are highlighted. (b) Tlr1 and Tlr6 expression in the three animals per strain of GSE248033.

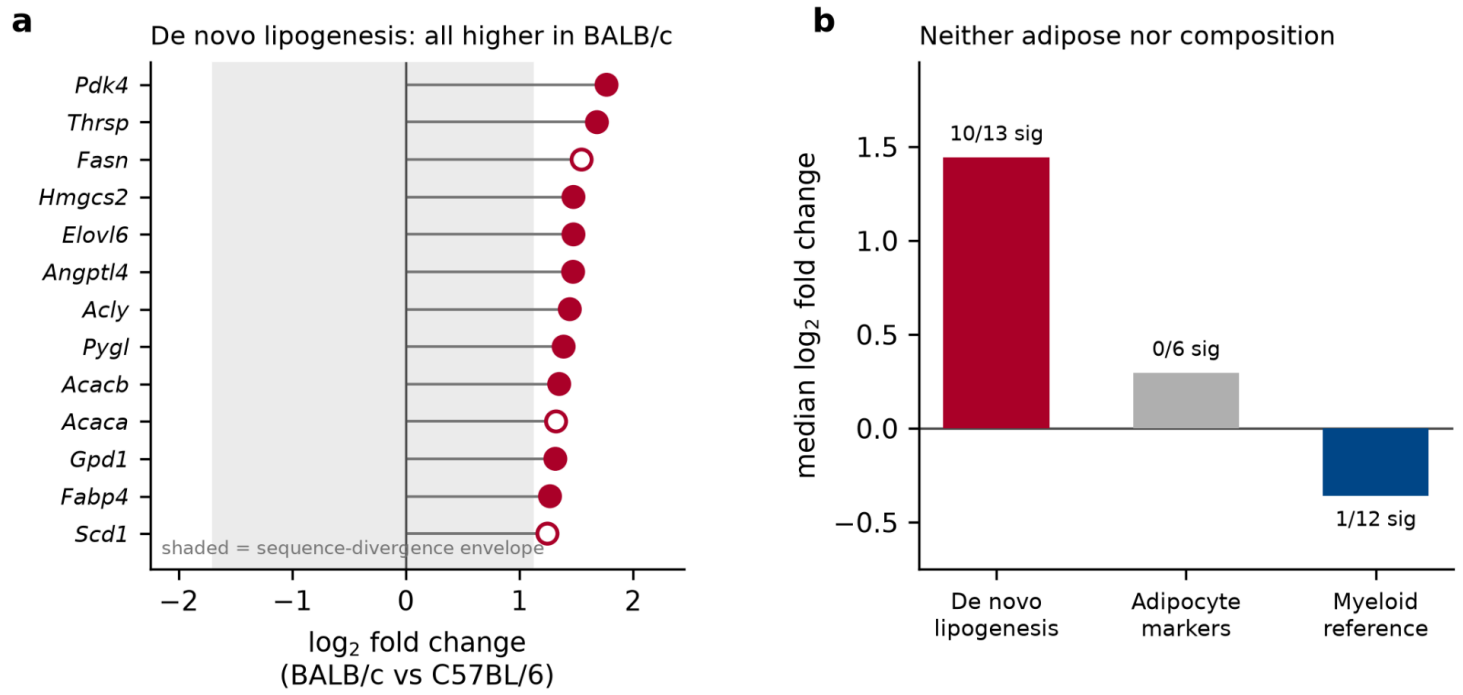


Figure 3

A lipid-synthesis module is the largest difference in the dataset but is not constitutive.

(a) Log2 fold change for 13 de novo lipogenesis and fatty-acid handling transcripts, BALB/c against C57BL/6; filled points reach a 5% false-discovery rate, open points do not. The shaded band marks the sequence-divergence envelope (-1.69 to +1.11) within which differential read mapping cannot be excluded; all ten significant members lie above it, the direction opposite to the one mapping bias produces. (b) Median log2 fold change for the module against two controls: adipocyte-restricted transcripts, which would move with the module if adipose tissue had been included in the dissection, and the myeloid reference panel, which indexes the leukocyte-fraction difference. Neither control moves with the module.