



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Mixed-species RNA-seq for elucidating non-cell-autonomous control of gene transcription

Citation for published version:

Qiu, J, Dando, O, Baxter, P, Hasel, P, Heron, S, Simpson, T & Hardingham, G 2018, 'Mixed-species RNA-seq for elucidating non-cell-autonomous control of gene transcription', *Nature Protocols*.
<https://doi.org/10.1038/s41596-018-0029-2>

Digital Object Identifier (DOI):

[10.1038/s41596-018-0029-2](https://doi.org/10.1038/s41596-018-0029-2)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Nature Protocols

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Mixed-species RNA-seq for elucidating non-cell-autonomous control of gene transcription

Jing Qiu^{1,2, #}, Owen Dando^{1,2,3,4, #*}, Paul Baxter^{1,2}, Philip Hasel^{1,2}, Samuel Heron⁵, T. Ian Simpson⁵,
and Giles E. Hardingham^{1,2*}

¹UK Dementia Research Institute at The University of Edinburgh, Edinburgh Medical School, 47 Little France Crescent, Edinburgh EH16 4TJ, UK

²Centre for Discovery Brain Sciences and ³Simons Initiative for the Developing Brain, Deanery of Biomedical Sciences, Edinburgh Medical School, University of Edinburgh, Edinburgh, EH8 9XD, UK.

⁴Centre for Brain Development and Repair, Institute for Stem Cell Biology and Regenerative Medicine, National Centre for Biological Sciences, Bangalore 560065, India.

⁵School of Informatics, University of Edinburgh, Edinburgh, EH8 9AB, UK

Equal contribution

*Correspondence to Giles.Hardingham@ed.ac.uk or Owen.Dando@ed.ac.uk

Keywords: transcription, RNA-seq, mixed-species, tissue culture, neurons, astrocytes, microglia, bioinformatics, differential gene expression

EDITORIAL SUMMARY: This protocol describes the co-culture of cells from multiple species and, following RNA-Seq, the separation of reads according to species using the Sargasso bioinformatics pipeline in order to elucidate the effects of one cell type on the transcriptome of the others

TWEET: **Mixed species cell co-culture, RNA-Seq and read separation by species to elucidate non-cell-autonomous transcriptional control. @UKDRI**

Hasel, P. et al. Neurons and neuronal activity control gene expression in astrocytes to regulate their development and metabolism. Nat Commun 8, 15132, (2017).

Qiu, J. et al. Evidence for evolutionary divergence of activity-dependent gene expression in developing neurons. eLife. 5, e20337 (2016).

Abstract

Transcriptomic changes induced in one cell type by another mediate many biological processes in the brain and elsewhere; however, achieving artefact-free physical separation of cell types to study them is challenging and generally only allows for analysis of a single cell type. We describe an approach employing co-culture of distinct cell-types from different species, which enables physical cell sorting to be replaced by *in silico* RNA sequencing (RNA-seq) read sorting due to evolutionary divergence of mRNA sequence. As an exemplary experiment, we describe the co-culture of purified neurons, astrocytes, and microglia from different species (12–14 days). Following conventional RNA-seq, we then describe how to use our Python tool *Sargasso* (<http://statbio.github.io/Sargasso/>) to separate reads according to species and how to eliminate any artefacts borne out of imperfect genome annotation (10 hours). We show how this procedure, which requires no special skills beyond those that might normally be expected of wet-lab and bioinformatics researchers, enables the simultaneous transcriptomic profiling of different cell types, revealing the distinct influence of microglia on astrocytic and neuronal transcriptomes under inflammatory conditions.

INTRODUCTION

Signaling between different cell types is central to the development and function of multicellular organisms. The secretion of factors such as cytokines, growth factors, and steroids enable paracrine and hormonal signaling to take place. Additionally, membrane-anchored ligands, such as ephrins, semaphorins, notch ligands, and several EGF, interleukin and TNF family members, mediate contact-dependent juxtacrine signaling¹. Intracellular signals transduced by the activated receptors of these ligands frequently induce changes in gene transcription that mediate the physiological or pathological response to that signal. Thus, understanding non-cell-autonomous influences on transcriptomes can illuminate diverse physiological processes. Furthermore, since the influences between two cell types are rarely unidirectional, it is valuable to be able to profile different cell types' transcriptomes simultaneously. This is a challenging undertaking using conventional techniques, which typically involve the dissociation of cultures or tissues into a single cell suspension followed by sequential separation of distinct cell types. Not only can physical dissociation and sorting induce gene expression artefacts, but separation requires distinct cell surface markers on each cell type and specific antibodies to them to enable their labelling. Known markers may not exist and even if they do, physical cell sorting often suffers from a degree of impurity.

Development and overview of the protocol

In the brain, as in other tissues, multiple cell types co-exist with highly specialized functions, and reciprocal signaling among these cells plays a key role in the organ's development and function. For example, while neuronal activity controls many processes taking place in the neurons themselves ²⁻⁶, it can also influence the properties of astrocytes and oligodendrocytes ^{7,8}. Conversely, astrocytes can respond to external oxidative stress by providing protective support to nearby neurons ^{4,9-11}. Additionally, astrocytes signal to neurons to play key roles in neuronal development ^{12,13}. Microglia also signal to and interact with neurons, both in development and in diseased states, while reciprocal signals from neurons to microglia, via neuron-derived secreted or cell surface ligands, can have strong immunomodulatory effects on the microglia themselves ¹⁴⁻¹⁷. Moreover, the importance of signaling among more than two cell types was underlined by the recent discovery that pro-inflammatory microglia induce astrocytes to express a phenotype that is toxic to nearby neurons ¹⁸.

The physiological responses of one cell type to signals originating from a different cell type, as described above, involve gene expression changes, though the gene networks involved and the signals controlling them are often not fully understood. To improve our understanding of these mechanisms, the analysis of multiple cell types within multi-cell-type systems would be desirable. This is not a straightforward task to perform, since neurons, astrocytes, oligodendrocytes, and microglia are morphologically complex cells and create specialized intertwined structures that make their physical separation difficult. We developed an approach to enable the simultaneous profiling of distinct cell types without having to physically sort them. In the strategy we adopted, we first co-cultured distinct neural cell-types from closely-related species; we then implemented an RNA sequencing (RNA-seq) protocol, which was followed by a workflow through which we separated the reads *in silico* according to species ⁷. RNA-seq reads from the analysis of an unsorted co-culture of distinct cell types coming from the same species would not be able to be attributed to one cell type or another. However, if different species are used for the different cell types, then evolutionary divergence of homologous sequences at the nucleotide level, even if the difference amounts to a single nucleotide in a paired-end read, will enable the read to be attributed to a specific species, and, by extension, a specific cell type. Our approach requires that between distinct species, there is sufficient divergence in gene sequences that RNA-seq paired-end reads

(typically 50–100 nt each) can be distinguished by species in the majority of cases for the overwhelming majority of genes. We confirmed that this was the case for rats and mice, and used a system of mouse astrocytes, cultured in the presence or absence of rat neurons, to determine the influence exerted by neurons on the astrocytic transcriptome ⁷. In another application of the same system, we identified the influence of synaptic activity on astrocytic gene expression ⁷.

The Procedure detailed below comprises two main parts. The first consists of the set-up of a mixed-species cell co-culture. As an example, we describe the generation and validation of neurons, astrocytes, and microglia of mouse, human, and rat origin, respectively. The previously published use of this protocol involved the use of just two cell types ⁷ (neurons (rat) and astrocytes (mouse)). The inclusion of a third cell type, while modifying the steps of the procedure, involves no conceptual changes to the workflow, which can be applied equally to any number of cell types or species combinations (while adhering to guidelines outlines below). While these three cell types are cultured separately by standard procedures ¹⁹⁻²², we describe how they are combined to create co-cultures. We then describe how to maintain these cells in the presence or absence of an activator of microglia (lipopolysaccharide, LPS) to determine how neurons and astrocytes respond to the presence of unchallenged versus activated microglia, while simultaneously profiling the response of the microglia themselves. The first part of the protocol ends with the extraction of RNA from the cultures, whereupon conventional RNA-seq can either be performed in the laboratory or by an external facility, using standard procedures ²³.

The second part of the protocol describes how raw RNA-seq reads are attributed through a combination of mapping approaches to the genomes of all relevant species, followed by their assignment to specific species using the *Sargasso* (Sargasso Assigns Reads to Genomes According to Species-Specific Origin) tool. We have previously shown that the separation of rat and mouse reads is straightforward⁷, and have now confirmed that this is also the case for the three-way separation of mouse, rat and human reads, and that implementing the *Sargasso* pipeline does not skew absolute FPKM (fragments per kilobase of transcript per million mapped reads) values substantially in the vast majority of genes or transcripts (Supplementary Results, Supplementary Figure 1, Supplementary Figure 2).

In the example experiment discussed in the Anticipated Results, we go on to show how the post-*Sargasso*-pipeline species-separated reads can be fed directly to a standard read counting software package (e.g. *featureCounts*) and differential gene expression analysis tools (e.g. *DESeq2*). Again, in our previous publication⁷, when disambiguating between the mouse and rat genomes, we showed that the *Sargasso* workflow did not substantially affect differential gene expression analysis (see Supplementary Figure 1 c,d from Hasel *et al.* ⁷). We have now also confirmed that the *Sargasso* workflow does not substantially affect differential gene expression analysis in human–mouse–rat combined samples (Supplementary Results, Supplementary Figure 3).

Comparison with other methods

The characterization of one cell type's transcriptome in a mixture of several different cell types poses technical challenges associated with the need to physically separate cell types prior to transcriptional analysis. This physical separation is typically performed either by immuno-sorting or fluorescence-activated cell sorting (FACS), relying on cell type–specific cell surface markers or fluorescent protein expression, which is not always practical. Moreover, both techniques are subject to off-target cell-type contamination and loss of material from delicate subcellular regions, which is also an issue for another potential approach to cell-sorting, laser capture microdissection^{24,25}. In our previous publication⁷, we assessed the disadvantages of imperfect physical separation by simulating a sorting process that achieved 95% purity. We “contaminated” mouse astrocytic mRNA with mouse neuronal mRNA then performed RNA-seq, finding that 863 genes were more than two-fold more highly-expressed in the contaminated sample than in a sample of pure astrocytes, and the expression levels of 216 genes were more than ten-fold higher in the contaminated sample than in the mentioned pure ‘control’ sample⁷.

An additional issue to be aware of is the aberrant induction of transcriptional responses due to the sorting process itself: tissue dissociation, FACS, and immunopanning all induce strong immediate early gene expression, which can confound or occlude genuine data^{7,24,25}. For example, we found that the trypsin digestion required to create a single cell suspension before

implementing magnetic activated cell sorting (MACS) or FACS induced the expression of multiple genes, including *Fos*, *Fosb*, *Nr4a2*, and *Ier2*. Thus, the effects of a genuine signal or stimulation that results in the up-regulation of these genes could be occluded by the induction of these genes within the control sample due to the dissociation process. The full MACS and FACS protocols induced the aforementioned genes even further. For example, *Fosb* expression level increased over eight-fold as a consequence of FACS implementation and over 16-fold as a consequence of MACS implementation ⁷.

It is possible to circumvent the need for physical cell sorting by using the cell type-specific expression of tagged ribosomes like the EGFP-L10a ²⁶, with the caveat that this approach only identifies ribosome-associated mRNAs. Importantly, it should be pointed out that this technique, much like the physical sorting techniques, tends to be suitable only for the analysis of one cell type in any one sample. Mixed-species RNA-seq offers a complementary approach to the techniques described above ⁷. By avoiding the need to physically sort cells prior to RNA extraction, the artefacts and issues described above can be avoided. Moreover, there is no theoretical limit to the number of cell types that can be studied simultaneously.

We note also that other software tools, besides *Sargasso*, can be used to separate mammalian RNA-seq data according to the species of origin, such as *Disambiguate* ^{27,28}. However, these tools were primarily developed to study xenografts of human tumors in mouse hosts, so their use is limited to separating data originating from two species. They are characterized by a single in-built balance between sensitivity and specificity that may not be appropriate when the species being separated are closely related to each other or a different filtering strategy is required. For example *Disambiguate* implements a "permissive" approach where obtaining the maximum number of reads assigned per-species is prioritized over minimizing the mis-assignment of reads to the wrong species. *Sargasso*, on the other hand, enables RNA-seq data originating from more than two species to be effectively separated, even when those species are closely related. It also allows users to define the precision and recall level. Moreover, by utilizing multiple processors, when they are available, *Sargasso* also permits a fast execution of the analysis. Finally, it is user-

friendly — in most cases, mapping and separation of all RNA-seq data sets can be completed with a single command-line invocation.

Level of expertise needed

The establishment and characterization of the mixed-species co-cultures described requires no special skills. Library preparation for RNA-seq also relies on standard kits and protocols, although both library preparation and sequencing are typically performed in an institutional or commercial next-generation sequencing facility. Application of the *Sargasso* workflow to the raw RNA-seq reads requires no specific skills beyond those required for conventional RNA-seq read mapping and differential gene expression analysis. Thus, any researcher or bioinformatics service provider with experience in such analyses will be able to apply the *Sargasso* workflow to their RNA-seq reads.

Limitations

We have used the Procedure reported below to successfully disambiguate reads of human, mouse, and rat origin. As long as the core principles discussed in the Experimental design section below are adhered to, then in theory the protocol can be applied to any combination of species. However, particular combinations may require further considerations to be taken in account, particularly if species are close in evolutionary terms, leading to a substantial fraction of RNA-seq reads being discarded as ambiguous due to 100% cross-species conservation (Supplementary Results, Supplementary Figure 4). Of note, for closely related species, it is however possible to estimate the extent to which each gene will lose reads due to the disambiguation process, using theoretical read sets covering the species' transcriptome (e.g. see Supplementary Figure 1a of Hasel *et al.* ⁷). Moreover, it is even possible to apply a correcting factor to the post-*Sargasso* FPKM values to compensate for the aforementioned estimated read loss. However, for a simple 'off the shelf' protocol, we recommend an evolutionary gap of around 15 million years, if the accuracy of FPKM values themselves is of high importance.

More generally, for the data to be useful, the species' genomes need to have been sequenced to high accuracy, minimising sequencing errors or gaps, particularly within genes, otherwise spurious discrepancies between genome and read sequences may lead to reads being discarded, or even attributed to the wrong species. In addition, substantial differences in genome sequence or annotation quality between the species used may result in reads being incorrectly assigned. In our experiments we have found that the genomes of commonly used mammalian species (human, mouse and rat, cow and pig) to have been sequenced to suitable accuracy. If other species are being used with less complete genomes, the procedure below includes steps to determine which genes may be ignored due to incorrect assignment of reads arising from sequencing errors or gaps.

Factors relating to the RNA-sequencing procedure itself should also be assessed: the longer the reads that are sequenced, the more likely it is that they will span nucleotides that will permit the disambiguation between species-by covering a region that is not 100% conserved between the species; additionally, the sequencing depth required may be high if a minority cell type is very sparse. Of note, in our example protocol implementation described below we achieved data for the microglial transcriptome, which represents under 10% of the total cell number, by increasing overall read depth to 240 million.

A limitation of this approach is that cell types from different species will never co-exist in any natural physiological situation, raising the question of the validity of any results obtained. Thus, an additional consideration to be made while interpreting the data is whether any transcriptional changes induced in one cell type by another are due to their being derived from different animal species. There is a balance between obtaining biologically relevant results, through the use of species that are evolutionarily close to each other, while still permitting separation of RNA-seq reads. Although maintaining this balance requires that the particular choices of species are informed by biological knowledge of the systems being studied, controls involving single-species data can also be applied to test the likely extent of any issues. For example, in our previous publication⁷, when separating RNA-seq data originating from mouse and rat (median evolutionary separation time of 16 million years²⁹) we showed, with both simulated and real RNA-seq data, that

the great majority of mouse reads can be unambiguously assigned to mouse (see Supplementary Figure 1a, b from Hasel et al. ⁷). Thus, separation of reads was achieved in two mammalian species in which functional conservation and inter-species compatibility of many ligand–receptor pairs are thought to be high. However, we wanted to further test whether genes induced by rat neuronal activity in mouse astrocytes might be due to their different species of origin. We analyzed the expression of the activity-regulated astrocytic genes whose induction could be tracked in a single species co-culture by virtue of their expression being over ten times higher in astrocytes than in neurons (15 out of 56 genes). All genes were indeed induced in a single-species mouse neuron–astrocyte co-culture, but not in a neuronal monoculture, which means that, with this subset of genes, there was no evidence that having neurons from a different species results in erroneous effects of synaptic activity on astrocytic gene expression (see Supplementary Figure 6d from Hasel et al. ⁷). Other types of validation are possible for genes of particular interest, such as immunohistochemistry ³⁰ or translating ribosome affinity purification ²⁶.

A final caveat applies to any *in vitro* culture experiment: it is important to use the data to generate hypotheses as to the behaviour of cells *in vivo*. An example can be found in the paper where we first published this approach ⁷. In that study, we were able to show that astrocytes in culture can be pushed towards a more *in vivo*-like phenotype by co-cultured neurons, that non-cell-autonomous signaling maintains the expression of neuronally-regulated astrocytic genes *in vivo*, and, finally, that neuronally-controlled astrocytic genes selectively showed age- and disease-associated deregulation *in vivo* ⁷.

Applications

The Procedure described below is centered on investigation of the effect of non-cell-autonomous signaling on gene transcription among neurons, astrocytes, and microglia of mouse, human, and rat origin, respectively (Steps 1–22), in the presence and absence of a pro-inflammatory stimulus (LPS) which activates microglial toll-like receptor 4 ³¹. However, this particular stimulus could be replaced with any number of others designed to influence signaling in one or more of the aforementioned cell types, depending on the experimental question. Examples

could include interleukin-13 to promote an anti-inflammatory microglial phenotype, a high K⁺ solution to induce neuronal Ca²⁺ influx, or tumor necrosis factor to promote a reactive astrocyte phenotype^{18,19,32}.

Moreover, the mixed-species RNA-seq workflow can be applied to any cell type or species combination, provided it is tailored to the experimental question or the practicalities of obtaining the cells required. In the context of biomedical research involving mammalian systems, the majority of cells employed are of mouse, rat or human origin. The use of these species is recommended given their well annotated genomes. There is no *a priori* reason why cell types of other species cannot be used, assuming they are separated evolutionarily by 15 million years or more. The steps described in the procedure are applicable to all species combinations, and the controls described will ensure the data obtained by any species combination is reliable.

In terms of the cell types to be used, the *Sargasso* Procedure can be employed with any combination of cell types to suit the particular biological question. Clearly, a relevant biological question is likely to involve the combination of cell types that are juxtaposed in a normal physiological situation. For example, other types of mixed species co-culture experiments that are underway in our laboratory include the co-culture of human endothelial cells and mouse pericytes to study gene regulation associated with reciprocal signaling between these cell types. Moreover, the addition of rat microglia to these co-cultures in the presence or absence of inflammatory stimuli, is being employed to understand how signals associated with inflammatory insults influence the integrity of endothelial cell–pericyte mono-layers.

It is also important to note that application of the Procedure need not necessarily be restricted to *in vitro* preparations. We are also using this approach to probe the fate of xenotransplanted cells *in vivo*. The introduction of human wild-type and disease-associated induced pluripotent stem cell (iPSC)-derived neurons, astrocytes, and oligodendrocytes into the rodent brain is proving to be a valuable research tool³³⁻³⁵. We are using the *Sargasso* pipeline–based approach to profile both transplanted cells and ‘host tissue’ to understand the development and maturation of the transplanted cells, as well as the host’s response to the ‘foreign’ tissue.

We have also reported a different use of the *Sargasso* pipeline: to study the expression of both human and mouse orthologues of the same gene in the same cell ³². The cells used in this study (neurons) were obtained from the Tc1 mouse model of Down syndrome, in which mice carry a stably segregating extrachromosomal copy of human chromosome-21 ³⁶. By using the *Sargasso* pipeline, followed by standard read mapping and *DESeq2* differential gene expression analysis, we were able to study the activity-dependent regulation of human chromosome-21 genes, as well as their mouse orthologues, in the same cell, uncovering differences in the regulation of certain orthologues between human and mouse ³².

Other potential uses of mixed species co-culture and the *Sargasso* pipeline include the study of the horizontal transfer of mRNA or miRNA from one cell to another, such as via exosomes ³⁷. Finally, although outside the scope of the current protocol, use of the *Sargasso* pipeline is not restricted to the separation of RNA-seq reads, but it could be adapted to any nucleic acid sample studied by next-generation sequencing. A current area of development is to use mixed species co-culture, combined with chromatin immunoprecipitation sequencing (ChIP-seq) or short bisulfite sequencing to understand the influence of one cell type on another's epigenome, such as DNA methylation or histone modification.

Experimental design

Controls. Regardless of the biological question to be answered, several core principles should be adhered to ensure the robustness of the data collected with the present protocol, in order to draw reliable conclusions. Firstly, the cell types to be studied must be characterized, ideally by immunohistochemistry, in order to confirm presence of expected markers, and the absence of markers of potential contaminating cell types. As illustrated by the images reported in Fig. 1, the neurons, astrocytes, and microglia from our example experiment are positive for their identifying markers (Neuro-ChromTM, Gfap, Iba1, respectively), and negative for each other's markers. Secondly, reference RNA-seq data sets need to be generated (or be available from public databases), in which each cell type (or species) of interest is absent but all other cell types (or species) in the full co-culture are present. These data are desirable to have, to define the small

number of genes whose reads from one species are attributed to the incorrect species' genome, either due to imperfect sequencing within the reference genome, polymorphisms in the precise genetic background of the animal from which the cells were derived, or failures in short RNA-seq read mapping (Steps 30–34). Although this mis-attribution problem affects only a small number of genes, it can by definition occur in differentially expressed genes when analyzing the influence that the presence of one cell type has on another's transcriptome. Ideally, such data sets will be generated for each experimental condition (for example, for each dose of a drug applied); this approach is desirable as the set of genes expressed in each condition (and thus potentially the source of spurious read assignments) may not be fully encompassed by those expressed in the basal condition. We routinely disregard those genes where mis-attributed reads are estimated to be greater than 10% of the total number of reads. At this threshold, mis-attributed reads could falsely report an 11% induction of a gene's expression due to the presence of an additional cell type. In the procedure below, we describe the generation not only of the neuron/astrocyte/microglia co-cultures, but also the two-species co-cultures of microglia/astrocytes, microglia/neurons, and astrocytes/neurons, necessary to correct for any read mis-attribution. Their preparation follows the exact procedure for the triple culture, with the modification of simply leaving out one of the three cell types.

Sequencing depth. The final consideration to be made before implementing the Procedure is the depth of sequencing of the RNA-seq step (Step 23). Due consideration should be given to the desired depth with respect to the least abundant cell type in the co-culture mixture and/or the cell type contributing the least amount of RNA to the total mixed-species RNA pool, and the total sequencing depth should be scaled accordingly. The choice of the level of read depth should be made also taking into account that a proportion of reads will fail to be unambiguously mapped to any species, due to the fact that they are 100% conserved across two species or more, or because they do not reach one or more of the thresholds that have been imposed to achieve the required specificity of species assignment (for example, the number of base mismatches allowed between read and genome, or the proportion of the read that is allowed to be left unmapped). The proportion of these 'unmapped' reads will depend on the evolutionary closeness of the species

employed, and it can be determined by applying the *Sargasso* workflow to an RNA-seq data set from the individual species in question.

MATERIALS

Reagents

- Antibody (Gfap, mouse IgG21, 1:500, Cell Signaling cat. no.3076)
- Antibody (Iba1, rabbit IgG, 1:500, Wako cat. no.019-19741)
- Antibody (Neuro-Chrom, mouse IgG1, 1:200, Millipore cat. no. MAB2300X)
- Antibiotic-Antimycotic (100X) (ThermoFisher, cat. no.15240062)
- Astrocyte medium (Caltag Medsystems, cat. no.1801)
- B-27 supplement (Invitrogen, cat. no. 17504-044)
- Calcium chloride solution (1 M, Sigma, cat. no. 21115)
- Cytosine β -D-arabino- furanoside hydrochloride (AraC, Sigma, cat. no. C6645)
- D-(+)-Glucose solution 45% wt/vol (Sigma, cat. no. G8769)
- D-MEM (GIBCO, cat. no. 31966-021)
- 1XDPBS (GIBCO, cat. no.14190250)
- Ethanol (Absolute, Sigma, cat. no. 51916)
- Fetal Bovine Serum (Invitrogen, cat. no. 10108-165)
- Glycine (Sigma cat. no. G7126)
- HEPES buffer (1 M, Sigma, cat. no. H0887)
- Human primary astrocytes (Caltag Medsystems, cat. no.SC1800)
- Kynurenic acid (Sigma, cat. no. K3375) **!CAUTION** Handle with care because kynurenic acid readily causes skin, eye, and respiratory irritation.
- Laminin (1mg/ml, Sigma, cat. no. L5520)
- L-Cysteine (Sigma, cat. no.168149)
- L-Glutamine 200 mM solution (Sigma, cat. no. G7513)
- Lipopolysaccharide (LPS, Sigma, cat. no. L2630)
- Magnesium chloride (Sigma, cat. no.M8266)
- MEM (+Earle's, - L-Glutamine) (GIBCO, cat. no. 21090-022)
- Neurobasal-A-Medium (NBA) (GIBCO, cat. no. 10888-022)
- OptiMEM I (GIBCO, cat. no. 31985-047)
- Papain (Merck Millipore, cat. no. 5125)
- Paraformaldehyde (PFA) (Sigma, cat. no. 158127) **!CAUTION** PFA is toxic if inhaled and can be absorbed through the skin. Use PFA in a fume hood and wear appropriate laboratory protective equipment.
- Phenol Red Solution (0.5% vol/vol, Sigma. cat. no. P0290)
- Phosphate buffer saline (PBS tablet, Sigma, cat. no. P4417)
- Poly-D-Lysine hydrobromide (BD Bioscience, cat. no.354210)
- Potassium chloride (Sigma, cat. no. P9333)
- Potassium sulfate (Sigma, cat. no. P9458)
- Rat serum (Harlan SeraLab Ltd, cat. no. R-0100B)
- Sodium bicarbonate solution (7.5% wt/vol, Sigma, cat. no. S8761)
- Sodium chloride (5 M, Sigma, cat. no. S5150)
- Sodium hydroxide (Sigma, cat. no. 71687)
- Sodium pyruvate (Sigma, cat. no. S8636)
- Sodium sulfate (Sigma, cat. no. 1614807)
- Sucrose (Sigma, cat. no. S9378)
- Sterile water (Sigma, cat. no. w3500-500ml)
- Trypsin-EDTA (Sigma, cat. no. T3924)
- Mouse embryos (obtained from time-mated dams, from institutional animal facility or commercial supplier such as Harlan or Charles River)

!CAUTION: For experiments requiring animal tissue, appropriate national laws and institutional regulatory board guidelines must be followed. In the case of the experiment described here, all procedures were performed in compliance with the UK Animals (Scientific Procedures) Act 1986 (licence P1351480E) and institutional regulations and approved by University of Edinburgh Local Ethical Review Board.

- Rat pups (obtained from time-mated dams, from institutional animal facility or commercial supplier such as Harlan or Charles River)

!CAUTION: For experiments requiring animal tissue, appropriate national laws and institutional regulatory board guidelines must be followed. In the case of the experiment described here, all procedures were performed in compliance with the UK Animals (Scientific Procedures) Act 1986 (licence P1351480E) and institutional regulations, and approved by University of Edinburgh Local Ethical Review Board.

SDS (Special Diet Services) Rat food and Mouse RM1 diet (maintenance diet) from DBM Scotland.

Reagent setup

▲ **CRITICAL** All reagents should be set up in a sterile environment.

Cytosine β -D-arabinofuranoside (AraC, 240 mM stock). Dissolve 67.12 mg of AraC in 1 ml of sterile water. Store the resulting solution in 10 μ l aliquots at -20 °C for up to 6 months. Freshly dilute these 10- μ l aliquots in 500 ml of NBA medium to prepare a working concentration of 5.5 μ M.

Human astrocytes culture. Purchase primary human astrocytes at passage one then expand and maintain them following the manufacturer's instructions.

Magnesium chloride solution (2 M stock). Dissolve 38.08 g of magnesium chloride in 200 ml of sterile water and store the resulting solution at 4 °C for up to 6 months.

Potassium sulfate solution (0.25 M stock). Dissolve 8.712 g of potassium sulfate in 200 ml of sterile water and store the resulting solution at 4 °C for up to 6 months.

Potassium chloride solution (3 M stock). Dissolve 44.73 g of potassium chloride in 200 ml of sterile water and store the resulting solution at 4 °C for up to 6 months.

Sodium hydroxide solution (1 M stock). Dissolve 4.0 g of sodium hydroxide in 100 ml of sterile water and store the resulting solution at 4 °C for up to 6 months.

Sodium sulfate solution (1 M stock). Dissolve 28.41 g of sodium sulfate in 200 ml of sterile water and store the resulting solution at 4 °C for up to 6 months.

Dissociation medium for neuron and microglia culture (DM, 500 ml). Add 40.9 ml of sodium sulfate solution (1M), 60 ml of potassium sulfate solution (0.25M), 1.46 ml of magnesium chloride solution (2M), 0.126 ml of calcium chloride solution (1M), 0.5 ml of HEPES buffer (1 M), 1 ml of Phenol Red solution, and 4 ml of D-(+)-Glucose Solution (45% wt/vol), and bring the total volume to 500 ml with sterile water. Filter the resulting mixture through a 0.22 μ m filter and store the filtrate at 4 °C for up to 1 month.

70% Ethanol (1000 ml). Mix 700 ml ethanol (absolute) with 300 ml of sterile water and store at RT.

Kynurenic acid (5mM stock, 160 ml). Add 158.56 mg of kynurenic acid to 5 ml of sterile water in a 50 ml Falcon tube and add to the tube 0.4 ml of Phenol Red solution. Gradually add a total of 500 μ l in ~100 μ l aliquots of the sodium hydroxide 1M solution and vortex extensively the Falcon tube between aliquot additions, until the kynurenic acid is fully dissolved. Add 0.4 ml of HEPES buffer (1M) and 4 ml of magnesium chloride 2 M solution and vortex. Transfer the resulting mixture to a 200-ml sterile bottle and adjust its volume to 160 ml using sterile water. Filter the mixture through a 0.22 μ m filter and store the filtrate in 10 ml aliquots at -20 °C for up to 6 months.

▲ **CRITICAL** It takes time for kynurenic acid to dissolve completely. Be sure to add sodium hydroxide gradually, around 100 µl at a time, and vortex extensively. If the solution turns pink, do not add acid to adjust the pH, but discard and start again.

Dissection solution (100 ml). Pool together 36 ml of DM and 4 ml of kynurenic acid stock. Use 0.2M sodium hydroxide solution to adjust the pH until the solution turns to rose-red in color. The dissection solution should be made up fresh each time.

D-MEM medium. Add 50 ml of fetal bovine serum and 5.5 ml of Antibiotic-Antimycotic to 500 ml of D-MEM. Sterile-filter using 500ml filter cups and store the filtrate at 4 °C for up to 7 d.

L-Glutamine (200 mM stock). Prepare 5 ml aliquots and store them at -20 °C for up to 6 months.

LPS (5 mg/ml stock). Dissolve 10 mg of LPS in 2 ml of sterile water. Store the resulting solution in 50-µl aliquots at -20 °C for up to 6 months.

1% (vol/vol) Neurobasal-A-Medium (1%NBA). Add 2 ml of Antibiotic-Antimycotic, 2 ml of rat serum, 4 ml of B-27 supplement, and 1 ml of L-glutamine stock solution to 191 ml of NBA. Sterile-filter through a 0.22-µm filter and store the filtrate at 4 °C for up to 7 d.

0% Neurobasal-A-Medium (0%NBA). Add 2 ml of Antibiotic-Antimycotic, 4 ml of B-27 supplement, and 1 ml of L-glutamine stock solution to 193 ml of NBA. Sterile-filter and store the filtrate at 4 °C for up to 7 d.

OptiMEM I+ medium. Add 4 ml of D-(+)-Glucose Solution (45% wt/vol) and 5 ml of Antibiotic-Antimycotic (100X) to 500 ml of OptiMEM I medium. Store the resulting solution at 4 °C for up to 7 d.

PFA (4% (wt/vol), 1000 ml). In a fume hood, add 40 g of sucrose to 1 l of PBS in a glass bottle. Add a magnetic stir bar to the bottle, then heat the bottle on a hot plate at 55–60 °C and be sure to keep the lid off. With the fume hood vent switched off, carefully measure out 40 g of PFA and add them into the heated solution, while stirring its contents, turning hood back on. Once the PFA is fully dissolved, allow the mixture to cool down to room temperature (~18 °C), remove any undissolved crystals by pouring through filter paper, and store 50-ml aliquots of the filtrate at -20 °C for up to 1 year.

!CAUTION: PFA is toxic, all the preparations should be performed in the fume hood.

▲ **CRITICAL** PFA powder is added when the solution has been heated to 55–60 °C. Make sure the temperature does not exceed 65 °C.

Papain dissociation solution (36,000 USP units ml⁻¹ stock). Pool together 10 ml of kynurenic acid stock, 90 ml of dissociation medium, and 45 mg of L-cysteine. Add to the resulting mixture sodium hydroxide solution (0.2 M) until the solution turns rose-red in color. Add to the red solution 120 mg of papain powder, mix well and filter through a 0.22 µm filter, and store the filtrate in 10-ml aliquots at -20 °C for up to 6 months.

Poly-D-Lysine hydrobromide (1mg ml⁻¹ stock). Dissolve 50 mg of Poly-D-Lysine in 50 ml of sterile water. Store the resulting solution in 5-ml aliquots at -20 °C for up to 6 months; once an aliquot has been thawed, store it at 4 °C and use it within 1 week.

SGG medium (500 ml). Add 11.4 ml of sodium chloride solution (5 M), 14.6 ml of sodium bicarbonate solution (7.5% wt/vol), 0.882 ml of potassium chloride solution (3M), 0.251 ml of magnesium chloride solution (2M), 1 ml of calcium chloride solution (1 M), 5 ml of HEPES buffer (1 M), 0.5 ml of glycine (1 M), 6 ml of D-(+)-glucose solution (45% wt/vol), 2.5 ml of sodium pyruvate solution (0.1M), and 1 ml of Phenol Red (0.5 % vol/vol). Store the resulting solution at 4 °C for up to 7 d.

TMo medium (500 ml). Pool together 445 ml of SGG medium, 50 ml of MEM (+Earle's, - Glutamine), and 5 ml of Antibiotic-Antimycotic. Store the resulting solution at 4 °C for up to 7 d.

Animal housing and maintenance. The pregnant mice and rats are housed with other female animals which become pregnant on the same day and housed individually a few days before the litters are born. The cages are checked twice a day and the beddings are changed weekly. The

food are SDS (Special Diet Services) Rat and Mouse RM1 diet (maintenance diet) from DBM Scotland. The enrichment includes tunnels, nesting and chew sticks.

Equipment

▲ **CRITICAL** Note that a bioinformatics researcher or service provider may have several, or all, the required software and input file dependencies for the *Sargasso* pipeline already installed.

- Adhesive matt (INFORS HT, cat. no. 63240)
- Agilent 2100 Bioanalyzer (Agilent, cat. no. G2939A)
- Aluminum foil
- Centrifuge tubes, 15 ml (Greiner, cat. no. 188271)
- Centrifuge tubes, 50 ml (Greiner, cat. no. 227261)
- Dissection microscope
- Falcon tube 50ml (Fisher scientific, cat. no. 14-432-22)
- Filtered cups (Steritop, Millipore, cat. no. SCGPT05RE)
- Hemocytometer (Hawksley, cat. no. AC1000)
- Humidified tissue culture incubator (37 °C, 5% CO₂, 20% O₂)
- Mr. Frosty™ freezing container (Thermo Scientific, cat. no. 5100-0001)
- Orbital shaker (Fisher Scientific, cat. no. SGM-250-030K)
- Petri dishes, 140 mm (Greiner, cat. no. 639102)
- Petri dishes, 35 mm x 10 mm (Greiner, cat. no. 627160)
- Polystyrene round-bottom tubes (Fisher, cat. no. TKV-180-110X)
- Pipettes, 5 ml (Greiner, cat. no. 710180)
- Pipettes, 10 ml (Greiner, cat. no. 607180)
- Pipettes, 25 ml (Greiner, cat. no. 760180)
- RNA 6000 Nano Chips (Agilent, cat. no. 400753)
- RNA Isolation Kit (Roche, cat. no. 11828665001)
- Spring scissors - 8mm blades (FST, cat. no. 15024-10)
- Tissue culture plates, 24-well, with lid (Greiner, cat. no. 662160)
- Tissue culture flask, 75 cm² (T75, Greiner, cat. no. 658170)
- Cages (Techniplast, 1145T for mice: 369x156x132; 1290T for rats: 425x265x155 mm).
- Unplugged Pasteur pipettes, 230 mm (Fisher Scientific, cat. no. FB50253)
- Vacuum filter (0.22 µm) unit, 500 ml (Fisher Scientific, cat. no. FDR-120-090W)

Hardware

- Operating system: Users of the protocol must have access to a Unix-like operating system, such as Linux or macOS, with a shell environment in which to execute commands, and the Python programming language (<https://www.python.org>) and Python package management system, *pip* (<https://pip.pypa.io/>), installed.

CRITICAL: All software was run on Ubuntu Linux 14.04.5 LTS

CRITICAL: The computing hardware used must be of sufficiently high specifications to run all stages of the protocol; note, for example, that the STAR read aligner requires approximately 30 Gb of RAM.

CRITICAL: Unless otherwise stated, the commands listed in the protocol should be run within a terminal window.

▲ **CRITICAL** Throughout this protocol, commands are indicated in Courier fonts, preceded by a ">" sign. Courier fonts are also used to indicate text file contents.

Software:

CRITICAL Since software tools are frequently updated, we suggest consulting the user manuals for compatibility; however, we do not foresee that any particular problems should arise if newer versions of these tools are used.

- Sargasso (<https://github.com/statbio/Sargasso>, doi:10.5281/zenodo.260123, also available as Sargasso-1.1.zip in Supplementary Software 1)
- short RNA-seq read aligner *STAR* (<https://github.com/alexdobin/STAR>)³⁸
- NGS alignment processing tool *sambamba* (<https://github.com/biod/sambamba>)³⁹.
CRITICAL We obtained the results detailed in this protocol using version 1.1 of *Sargasso* (using version 2.7.6 of Python, version 2.4.0i of *STAR* and version v0.5.4 of *sambamba*).
 - version 1.5.1 of *featureCounts* (<http://subread.sourceforge.net>)
 - version 1.16.1 of *DESeq2* (<https://bioconductor.org/packages/release/bioc/html/DESeq2.html>).
 - version 0.8.2 of *Salmon* (<https://combine-lab.github.io/salmon/>)⁴⁰,
 - version 1.4.0 of *tximport* (<https://bioconductor.org/packages/release/bioc/html/tximport.html>)⁴¹

Equipment setup

Poly-D-Lysine-coated 24-well plates or T75 flasks. In a sterile hood, prepare a coating mix by adding 1.33 ml of Poly-D-Lysine stock into 100 ml of sterile water to prepare a working solution of 0.013 mg.ml⁻¹. Add 0.4 ml of coating mix to each well of a 24-well tissue culture plate, or 10 ml of coating solution to each T75 flask. Incubate the plates or flasks for 2 h at 37 °C. Aspirate the coating mix from the plates or flasks, rinse the plates or flask with sterile water, aspirate the water, and leave the plates or flasks in a sterile hood with the lid off until they are completely dry (~1 h).

▲ **CRITICAL** Be sure to wash the plates or flasks with sterile water rather than with PBS, as the latter forms crystals when it dries.

Install *STAR* and *sambamba*. Download *STAR* from <https://github.com/alexdobin/STAR>, and *sambamba* from <https://github.com/biod/sambamba>. In both cases, precompiled executables are available for Linux and macOS.

Download and extract genome sequence and annotation files. Directions on downloading genome information are given in Box 1.

Download raw RNA-seq data to server. Raw RNA-seq data will have been provided by your preferred sequencing facility. Alternatively, existing data sets can be downloaded from an online database such as the NCBI Gene Expression Omnibus (<https://www.ncbi.nlm.nih.gov/geo/>) or the European Nucleotide Archive (<https://www.ebi.ac.uk/ena>). The data used in the Anticipated Results section of this protocol can be downloaded from the European Nucleotide Archive (accession number E-MTAB-5987).

Input file formats: The main input files supplied to *Sargasso* should be raw per-sample sequencing reads, as produced by a sequencing facility, in FASTQ format. These files are also assumed to have been *gzip*-compressed. *Sargasso* also requires genome sequence and annotation files for the species being separated. Genome sequences should be supplied in FASTA format; multiple reference sequences can be contained in the same FASTA file, or separate FASTA files can be provided for each reference sequence. Gene annotations should be supplied in GTF format. Refer to the *STAR* manual regarding requirements for these files; in particular, the chromosome names in the annotation files must match those in the genome sequence files. In the exemplary application of this protocol below, genome sequences and annotations from Ensembl version 86 (ref. ⁴²) were used.

Install *virtualenv* and *virtualenvwrapper*. *Sargasso* has a number of dependencies on other Python packages, and hence we strongly recommend (though it is not required) to install it in an isolated Python environment using the *virtualenv* (<https://virtualenv.pypa.io>) tool. In addition, the

virtualenvwrapper (<https://virtualenvwrapper.readthedocs.io>) tool helps to manage multiple virtual environments. Install both tools using the Python package management system, *pip*:

```
> pip install virtualenv
> pip install virtualenvwrapper
```

We recommend following the *virtualenvwrapper* installation documentation to ensure *virtualenvwrapper* functions correctly with the user's particular command-line shell (e.g. Bash).

Install *Sargasso* in a new Python virtual environment. Create a new virtual environment project directory called “neurons_astrocytes_microglia” (or any other appropriate name of the user's choice), and then install *Sargasso* using *pip*:

```
> mkproject neurons_astrocytes_microglia
> pip install git+https://github.com/statbio/sargasso.git
```

Box 1 | Download and extract genome sequence and annotation files

In this example, we will use genome sequences and annotation files downloaded from Ensembl ⁴², however, other sources of sequence and annotation files, such as the UCSC Genome Browser, would work equally well.

1 | Download the sequence and annotation files for human, mouse, and rat, and place them in an appropriate location in your directory tree. Please note that here, we assume files are stored within the home directory of a user called “sargasso_user”:

```
> wget ftp://ftp.ensembl.org/pub/release-86/fasta/homo_sapiens/dna/Homo_sapiens.GRCh38.dna.primary_assembly.fa.gz
> gunzip Homo_sapiens.GRCh38.dna.primary_assembly.fa.gz
> wget ftp://ftp.ensembl.org/pub/release-86/gtf/homo_sapiens/Homo_sapiens.GRCh38.86.gtf.gz
> gunzip Homo_sapiens.GRCh38.86.gtf.gz
> wget ftp://ftp.ensembl.org/pub/release-86/fasta/mus_musculus/dna/Mus_musculus.GRCm38.dna.primary_assembly.fa.gz
> gunzip Mus_musculus.GRCm38.dna.primary_assembly.fa.gz
> wget ftp://ftp.ensembl.org/pub/release-86/gtf/mus_musculus/Mus_musculus.GRCm38.86.gtf.gz
> gunzip Mus_musculus.GRCm38.86.gtf.gz
> wget ftp://ftp.ensembl.org/pub/release-86/fasta/rattus_norvegicus/dna/Rattus_norvegicus.Rnor_6.0.dna.toplevel.fa.gz
> gunzip Rattus_norvegicus.Rnor_6.0.dna.toplevel.fa.gz
> wget ftp://ftp.ensembl.org/pub/release-86/gtf/rattus_norvegicus/Rattus_norvegicus.Rnor_6.0.86.gtf.gz
```

```
> gunzip Rattus_norvegicus.Rnor_6.0.86.gtf.gz
> mkdir /home/sargasso_user/sargasso_data
> mv *.gtf /home/sargasso_user/sargasso_data
```

2 | Place the genome FASTA file(s) in per-species genome directories:

```
> mkdir /home/sargasso_user/sargasso_data/human_primary_assembly
> mv Homo_sapiens.GRCh38.dna.primary_assembly.fa
/home/sargasso_user/sargasso_data/human_primary_assembly
> mkdir /home/sargasso_user/sargasso_data/mouse_primary_assembly
> mv Mus_musculus.GRCm38.dna.primary_assembly.fa
/home/sargasso_user/sargasso_data/mouse_primary_assembly
> mkdir /home/sargasso_user/sargasso_data/rat_toplevel
> mv Rattus_norvegicus.Rnor_6.0.dna.toplevel.fa
/home/sargasso_user/sargasso_data/rat_toplevel
```

? TROUBLESHOOTING

PROCEDURE

CRITICAL: The culture of microglia (Steps 1-9), astrocytes (Steps 10-15) and neurons (Steps 16-18) described sequentially below, have overlapping timelines. Throughout the Procedure, each day is numbered relative to the first day of Step 1.

Rat microglia culture •TIMING 12–14 d

1 | On day 1, take 10–12 rat pups of postnatal day 0–2, dislocate necks with forceps and decapitate with scissors. This number is needed to reliably produce the microglia required, and is kept to a minimum according to ARRIVE principles. Keep the heads in a petri dish containing freshly prepared dissection solution. Stabilize the head with a pair of fine needles and make a horizontal slit in the skin with fine scissors, then cut up along the midline. Cut the skull along the midline while pulling up with scissors; make horizontal incisions on each side, at the bottom and at the top. Open up the skull to reveal the brain using forceps, scoop out the brain from underneath and then leave it in a petri dish containing dissection solution. Under a dissection microscope, cut off the olfactory bulbs, and using forceps peel the cortical hemispheres away from the hindbrain and thalamus.

▲ **CRITICAL STEP** To preserve the brains, make sure they are fully submerged in dissection solution and agitate the petri dish occasionally to equilibrate the medium.

!CAUTION: For experiments requiring animal tissue, appropriate national laws and institutional regulatory board guidelines must be followed. In the case of the experiment described here, all

procedures were performed in compliance with the UK Animals (Scientific Procedures) Act 1986 (licence P1351480E) and institutional regulations, and approved by University of Edinburgh Local Ethical Review Board.

2 | To begin the enzymatic dissociation, warm up the papain dissociation solution in a 37 °C water bath. Transfer 6–8 cortical hemispheres into one polystyrene round-bottom tube and add to each tube 2 ml of papain dissociation solution. Leave the tubes for 20 min in a 37 °C water bath and carefully shake them every 5 min. After 20 min incubation, aspirate the solution and replace it with fresh papain dissociation solution. Incubate for another 20 min and carefully shake the tubes every 5 min.

3 | Remove the papain dissociation solution and wash the brains twice with 2 ml of dissection solution. Between each washing, hold each tube horizontally and rotate slowly whilst rocking it back and forth to wash the enzyme off the walls and agitate meninges away from the cortical hemispheres. Remove the dissection solution, wash the cortical hemispheres twice with 2 ml of pre-warmed (37 °C) 1%NBA, and leave in pre-warmed (37 °C) 1%NBA medium to proceed to Step 4.

4 | Take two 15-ml Falcon tubes (label them A and B) for every 6–8 cortical hemispheres and add 10 ml of 1%NBA medium to tube A. Remove the NBA solution from the polystyrene round-bottom tube that contains the cortical hemispheres from Step 3, and using a 5 ml pipette, take 2 ml of 1% NBA medium from tube A and add to the brain tissues contained in the polystyrene round-bottom tube. Pipette up and down quickly 50–60 times (taking care not to let the pipette tip touch the bottom of the tube), and add another 2 ml of 1%NBA from tube A. Leave the tube to rest for 2 min for all the debris to settle. Single cells will remain in the supernatant.

5 | Repeat Step 4 for all the other cortical hemisphere samples.

6 | Remove the supernatant (4 ml) from the polystyrene round-bottom tubes very gently without touching the tissue at the bottom of the tube and transfer it into one of the empty 15 ml Falcon tubes (tube B) from Step 4.

7 | Add another 4 ml of 1%NBA from tube A to each polystyrene round-bottom tube (which leaves 2 ml of 1%NBA in tube A), pipette up and down quickly 50–60 times (taking care not to let the pipette tip touch the bottom of the tube), and leave the tubes to rest for 2 min for all the debris to

settle. Remove the supernatant (4 ml) from the polystyrene round-bottom tubes very gently without touching the tissue at the bottom of the tube and transfer it into Tube B. Take the remaining 2 ml of 1%NBA and dissociate the remaining tissue hard, pipetting up and down 50-60 times with the pipette tip pushed to the bottom of the tube. Leave to settle for 5 min and transfer this last 2 ml of cell suspension to tube B (making 10ml in total).

8 | To plate, collect the homogenized cell suspension from all the 'B' labelled 15 ml Falcon tubes prepared in Steps 6 and 7 and transfer it to a single sterile bottle. Adjust the volume of this mixture to 100 ml using pre-warmed (37 °C) OptiMEM I+ medium. Mix by pipetting up and down with a 25 ml pipette. Plate 10 ml of the cell suspension into a poly-D-Lysine-coated T75 flask. Allow the cells to settle for 2.5 h in an incubator (37 °C, 5% CO₂)

▲ **CRITICAL STEP** Be sure the cells are mixed well before plating them. For this purpose, homogenize the diluted mixture by pipetting up and down with a 25-ml pipette quickly, but avoid excessive swirling, which causes an uneven distribution of cells in solution.

9 | Aspirate the OptiMEM I+ medium, gently add 10 ml of pre-warmed (37 °C) D-MEM medium into each flask, and leave the flask in the incubator for the cells to grow for 12 d, until their use in Step 19 (there is some flexibility here-12-14 days is acceptable). Replace the medium with pre-warmed (37 °C) fresh D-MEM medium every 3–4 d.

Human astrocyte culture •TIMING 3–4 d

10 | On day 2, i.e. the day after the start of microglia culture (Step 1), take a vial containing around 1 million frozen primary human astrocytes stored at -80 °C and thaw its contents rapidly in a 37 °C water bath. Spray the vial with 70% ethanol and, using a 5-ml plastic pipette, gently transfer cells to a poly-D-lysine-coated T75 flask containing 15 ml of pre-warmed (37 °C) astrocyte medium. Allow cells to settle for at least 12 h.

? TROUBLESHOOTING

11 | On day 3, aspirate the medium from the T75 flask, gently wash the bottom of the flask with 10 ml of 1XDPBS, aspirate the DPBS, add to the flask 4–5 ml of pre-warmed (37 °C) trypsin-EDTA.

▲ **CRITICAL STEP** Be sure to wash the residual astrocyte medium away with DPBS, as the serum present in DPBS medium can inhibit trypsin activity.

12 | Place the flask in a 37 °C incubator for 3–5 min or until all cells have rounded up.

13 | Take the flask out of the incubator and gently tap its sides to force cells to detach from the flask surface. Check under a microscope to make sure all the astrocytes are detached. Add 10 ml of D-MEM medium to the flask and wash all remaining astrocytes from the surface.

14 | Use a pipette to transfer all the medium from the flask to a 15 ml Falcon tube, then centrifuge the tube at 150 *g* for 4 min at room temperature (~18 °C).

15 | Aspirate the supernatant from the Falcon tube, add 1 ml of astrocyte medium to the pellet and re-suspend the pellet using a 1-ml pipette. Count the cells and plate them at a density of 30,000–50,000 cells per well in astrocyte medium on a poly-D-Lysine-coated 24-well plate. The number of wells required is four per condition per culture type. Allocate astrocyte wells for generation of the 3-species co-culture, as well as the two species neuron/astrocyte and microglia/astrocyte co-cultures. Leave cells to grow for 1–3 d until they reach 80–90% confluency.

PAUSE POINT Human astrocytes can be passaged up to 4 times before co-culturing them with mouse neurons. To passage the cells, take a confluent T75 flask of astrocytes, follow Steps 11–14, and then **resuspend the cell pellet [AU: In which medium?]** and plate them onto three T75 flasks in astrocyte medium (20ml per flask)

▲ CRITICAL STEP Before each passage, use a standard tissue culture microscope under phase-contrast to check for any significant changes in human astrocyte morphology. **The key change visible [AU: Edit correct?]** in late-passage astrocytes is a loss of processes, which is why we recommend only 4 passages.

Mouse neuron co-culture with astrocytes •TIMING 7–8 d

16 | On day 5, take 2–3 mouse embryos of embryonic day 17.5, dissect and dissociate the cortical hemispheres as described in Steps 1–7. This number is needed to reliably produce the neurons required, and is kept to a minimum according to ARRIVE principles.

17 | Collect the homogenized cell suspension from the 15 ml Falcon tubes into a sterile bottle and adjust the suspension's volume to 28 ml per embryo using pre-warmed (37 °C) OptiMEM 1+ medium. Remove the medium from human astrocytes from Step 15, mix the neuron suspension well with a 25 ml pipette and plate 0.5 ml of the mouse neuron cell suspension into wells containing human astrocytes. Plate 4 wells per condition, per co-culture type, and allocate half of the wells to the 3-species co-culture (onto which microglia are added, see below) and half to the 2-species astrocyte/neuron co-culture (no microglia). Ensure a set of astrocyte-containing wells are left without neurons plated onto them for generation of the 2-species microglia/astrocyte co-culture. Additionally, plate neurons onto poly-D-lysine-coated 24 well plates (no astrocytes) for generation of the 2-species microglia/neuron co-culture, again at 4 wells per condition. The density of mouse

neurons should be $9-13 \times 10^4$ cells per well. Leave the plates for 2.5 h in a 37 °C tissue culture incubator.

? TROUBLESHOOTING

18| Aspirate the OptiMEM I+ medium from each well and replace it with 1 ml of 1%NBA with AraC for each well; incubate the plates at 37 °C for 4 d. Three days following the plating of the mouse neurons, add 1 ml of fresh 1%NBA with AraC to each well. Leave the plates to incubate for another 3–4 d at 37 °C.

? TROUBLESHOOTING

Rat microglia co-culture with neurons and astrocytes •TIMING 3–4 d

19| On day 13, tighten the caps of the T75 flasks from Step 9 if the flasks have no filters, or wrap parafilm around the caps if there are filters, in order to sustain the CO₂ content at the medium inside the flasks. Tape the flasks down to an orbital shaker and shake at 250 rpm for 1 h at 37 °C to detach the microglia. At the end of this shaking procedure, the microglia are contained in the medium.

? TROUBLESHOOTING

20| Remove the culture medium containing the microglia and transfer to a fresh tube, then centrifuge at 150 g for 5 min at room temperature. Remove the supernatant and re-suspend the pellet in 2 ml of NBA medium without serum (0% NBA). Count the cells with a hemocytometer. The yield is typically 5 to 15 million cells. Adjust the concentration to 2×10^5 cells/ml in 0%NBA. Remove the medium from neuron/astrocyte co-culture from Step 18 and plate 1 ml of the microglia suspension into each well that is intended for 3-species co-culture. Repeat the process with the astrocyte-only and neuron-only culture wells, to create the 2-species neuron/microglia and astrocyte/microglia cultures respectively, which are maintained under the same conditions as the 3-species co-culture [AU: Please specify what the conditions are for all the cultures to be maintained].

21| OPTIONAL: If desired, perform immunostaining for neurons, astrocytes or microglia with antibodies targeting appropriate cell-type markers, for example Neuro-chrome, Gfap or Iba1. This may be useful if the experimenter wishes to confirm the identity of their cell type, and rule out the presence of other cell types in their preparation. The co-cultures should be fixed for immunostaining 1-3 days after microglial plating, by replacing the medium with 300 µl of PFA per

well at room temperature for 10–20 min, and then washing three times with PBS, after which a standard immunostaining protocol should be followed ⁴³.

22 | To subject the cells to an inflammatory challenge, on day 16 (if the microglia are harvested and plated on day 13), replace the medium with TMO medium. Then on day 17, challenge the co-cultures with an inflammatory stimulus, for instance by adding 10 µl of LPS stock (5 mg/ml) to each well for a final concentration of 50 µg/ml.

▲ **CRITICAL STEP** In the case described, the challenge the cells were subjected to is inflammatory in nature; however, the cells may, of course, be subjected to other types of challenges (see Introduction).

23 | On day 18, harvest total RNA and test RNA integrity by standard methods. In our laboratory, we use the Roche RNA isolation Kit implementing the standard protocol, and we confirm the quality of total RNA (RIN) on RNA 6000 Nano Chips in the Agilent 2100 Bioanalyzer following the manufacturer's instructions. Then submit RNA to your preferred facility for conventional RNA-seq at an appropriate depth (see Experimental Design).

? TROUBLESHOOTING

In silico separation of RNA-seq reads with *Sargasso* • **TIMING ~7 h**

24 | Construct, in the virtual environment project directory, a tab-separated file listing the raw RNA-seq read data files for each sample in the experiment. Each line in the file should correspond to a single RNA-seq sample. In the case of paired-end reads, each line should contain three columns: a sample name, and then two comma-separated lists of the gzipped FASTQ files which contain, respectively, the first and second mates of the paired-end RNA-seq reads for that sample. In the case of single-end reads, each line should contain two columns: a sample name and a comma-separated list of the gzipped FASTQ files containing reads for that sample. For example, if our first sample has the name SAMPLE_A, the raw read FASTQ files for this sample are contained in the directory “~/sargasso_data/sample_a”, and there are three pairs of paired-end read files for this sample (denoted “sample_a1”, “sample_a2”, and “sample_a3”). The first line of our samples file, “samples.tsv”, would then read:

```
SAMPLE_A
    sample_a/sample_a1_1.fastq.gz,sample_a/sample_a2_1.fastq.gz,sample_
a/sample_a3_1.fastq.gz
    sample_a/sample_a1_2.fastq.gz,sample_a/sample_a2_2.fastq.gz,sample_
a/sample_a3_2.fastq.gz
```

25 | Run the *Sargasso* pipeline to map and separate the per-sample raw RNA-seq reads according to their true species of origin. As discussed previously, *Sargasso* enables the user to control the

balance between sensitivity and specificity, when separating reads, via a number of thresholds that the mapped reads must get past; these thresholds are controlled through the command line options `--mismatch-threshold`, `--minmatch-threshold`, `--multimap-threshold`, `--overhang-threshold`, and `--reject-multimaps`. *Sargasso* also provides a number of “pre-baked” filtering strategies, which specify particular values for these options, and, thus, different balances between sensitivity and specificity. As an exemplary application of this protocol, here we use a pre-baked “conservative” strategy, in which minimizing the number of reads mis-assigned to the wrong species takes foremost priority, rather than recall of the maximum number of reads per species. In the “neurons_astrocytes_microglia” virtual environment project directory previously created (See Equipment Setup) execute:

```
> species_separator --num-threads=16 --reads-base-dir=/home/sargasso_user/sargasso_data/ --conservative --run-separation samples.tsv sargasso_output human /home/sargasso_user/sargasso_data/Homo_sapiens.GRCh38.86.gtf,/home/sargasso_user/sargasso_data/human_primary_assembly mouse /home/sargasso_user/sargasso_data/Mus_musculus.GRCm38.86.gtf,/home/sargasso_user/sargasso_data/mouse_primary_assembly rat /home/sargasso_user/sargasso_data/Rattus_norvegicus.Rnor_6.0.86.gtf,/home/sargasso_user/sargasso_data/rat_toplevel
```

The `--num-threads=16` option indicates that 16 processor cores should be used at any stage of the separation procedure that can be multi-threaded (note that while the *Sargasso* pipeline can be run with any number of threads, increasing the processing cores used will reduce execution time). The `--reads-base-dir` option allows a base directory to be set for RNA-seq data input files, which can simplify construction of the samples file “samples.tsv”. The `--run-separation` option indicates that the separation pipeline will be executed immediately in the background via the `nohup` command; execution can, alternatively, be postponed until a later time by omitting this option, particularly if further fine-grained alterations of the separation procedure are necessary. Here, we have also assumed that the samples file “samples.tsv” has been written in the virtual environment project directory. The *Sargasso* output will be written in the directory “sargasso_output”, which will also be created in the project directory; pipeline progress can be monitored by examining the file “nohup.out” that is written into the “sargasso_output” directory.

▲ **CRITICAL STEP** Please note that here we have specified genome sequence and annotation files for each species; thus, in this case, the first step executed by the *Sargasso* pipeline will be to build *STAR* indices for each genome. In subsequent pipeline runs, these pre-built *STAR* indices can be re-used to reduce execution time.

Examining the *Sargasso* output •TIMING 30 min

26 | On completion of the *Sargasso* pipeline, ensure that the “sargasso_output” directory contains the following sub-directories: “star_indices”, which contains a *STAR* index directory for each

species' genome'; "raw_reads", which contains symbolic links to the *gzip*-ped FASTQ files for each sample; "mapped_reads", which contains a BAM (Binary Alignment Map) file for each sample and species describing the mapping of the RNA-seq reads in that sample to the species' genome (whatever the true species of origin of those reads is) — please note that this sub-directory also holds log files output by *STAR* containing statistics describing its performance; "sorted_reads", which contains a BAM file for each sample and species, where the mapped reads above have been sorted in name order; and "filtered_reads", which contains a BAM file for each sample and species, describing the mapping of the RNA-seq reads determined to have originated from that species, to that species' genome.

▲ **CRITICAL STEP** If *Sargasso* has been run with the `--run-separation` option, the "sargasso_output" directory will also contain a file, "nohup.out", recording the commands executed during pipeline operation and their terminal output. In the case that any of the directories listed above do not contain the expected files on completion of the pipeline, "nohup.out" can be examined to pinpoint the source of any errors.

? TROUBLESHOOTING

27 | Ensure that two further log files are written after completion of the *Sargasso* pipeline. In the "filtered_reads" directory, "overall_filtering_summary.txt" contains per-sample statistics describing the number of reads that were assigned to each genome, or were rejected as not belonging to that species, or were rejected as ambiguous. In the top-level output directory, "execution_record.txt" contains a record of the command line options that were passed to *Sargasso*, together with the date and time of execution.

Downstream analysis •TIMING varies according to the analysis performed

28 | For each RNA-seq sample, the *Sargasso* pipeline's output consists of per-species BAM files describing the mapping of reads determined to have originated from those species. These BAM files can then be utilized as the input of any standard post-mapping RNA-seq analysis. Notably, no peculiarities are expected to occur in such analyses due to the processing by *Sargasso*. One possible downstream analysis would be to determine the differential gene expression for a particular cell type. To perform this type of analysis, count the reads assigned to each gene in each sample for the appropriate species by using each BAM file as input for a read-counting tool like *featureCounts*⁴⁴.

29 | Use the per-sample read count files produced by *featureCounts* as output to detect genes whose expression has statistically significantly changed using tools such as *DESeq2*⁴⁵ or *edgeR*⁴⁶. Briefly, to calculate differential gene expression with *DESeq2*, install first the *DESeq2* package within the R programming environment (the *DESeq2* package is part of Bioconductor, an open-source repository for tools for the analysis of high-throughput genomic

data: <https://bioconductor.org>). Using the per-sample read count files produced by *featureCounts* as input, construct a table in R containing all per-gene read counts for all samples. Construct also a second table containing metadata describing the grouping of samples, most importantly indicating which experimental condition is distinctive of each sample. These tables, together with a model formula describing how gene expression is believed to depend on the experimental groupings of samples, form the input to the main *DESeq2* functions, which perform library size normalization of per-gene read counts across samples, estimation of per-gene count variances, and statistical tests for per-gene differences in counts between experimental conditions using the negative binomial distribution. Finally, for any experimental condition of interest, extract the lists of genes expressed at significantly different levels, together with the relevant p-values and fold changes in expression. Notably, a similar workflow would be used if, instead of the *DESeq2* package, the reader were to use the *edgeR* package or some other differential expression tool; for more details on such standard differential gene expression processing see, for example, Anders *et al.* ⁴⁷.

Avoiding misinterpretation of data •TIMING 2 h

▲ **CRITICAL** When separating mixed-species RNA-seq data according to true species of origin, it is difficult to avoid making a small number of incorrect species assignments, and it is important to ensure that these reads do not lead to spurious conclusions. The approaches needed to avoid such misinterpretation may vary according to the particular downstream analysis that has been performed; in the present sub-section of the Procedure, as an illustration, we will provide instructions on how to implement one such possible approach when the objective of the approach is to determine differential gene expression.

▲ **CRITICAL** Note that the calculations described below could be executed in any number of computational environments, for example, in an Excel spreadsheet or within the R programming environment.

30 | Use the read counts that have been calculated for each species contained in the 3-species co-culture samples, and also for each species in the 2-species co-culture samples that do not contain one particular cell type [AU: Edits correct?], to determine those small number of genes which should be disregarded due to mis-attribution of reads to the wrong species. We illustrate here the case of gene induction in microglia by LPS, in the presence of neurons and astrocytes; here we must avoid misidentifying increased expression of rat microglial genes due to the incorrect assignment of reads truly originating from mouse neuronal and human astrocytic genes, including from those mouse and human genes themselves induced by the presence of the perturbed microglia (similar steps should be carried out to avoid misidentifying mouse neuronal genes induced by perturbed microglia in the presence of human astrocytes, and human astrocytic genes

induced by perturbed microglia in the presence of mouse neurons). Firstly, calculate for each rat gene an approximate gene “expression” in samples containing only mouse and human biological material, measured in fragments per kilobase per million mapped reads, via the formula:

$$FPKM_{mm_hs_i} = \frac{10^9 \times c_i}{l_i \times R},$$

where c_i is the number of reads assigned to rat gene i in the mouse plus human samples, l_i is the length of the longest transcript of rat gene i , and R is the total number of reads assigned to all genes of all species in the mouse plus human samples. Please note that if *featureCounts* has been used to count the reads assigned to genomic features, these read numbers can easily be extracted from the count summary files output.

31 | For each rat gene, and for each 3-species sample, calculate an approximate level of gene expression in the sample, measured in fragments per kilobase per million mapped reads (FPKM), via the same formula reported in Step 30:

$$FPKM_{rn_i} = \frac{10^9 \times c_i}{l_i \times R},$$

where c_i is now the number of reads assigned to the rat gene i in the 3-species sample (after the sample has undergone species separation), and R is now the total number of reads assigned to all rat genes in the sample.

32 | For each 3-species sample, calculate the approximate ratio of “mouse plus human” to rat RNA in the sample. To perform this task, examine the *Sargasso* log file, “overall_filtering_summary.txt”. The fields with headers reading “Filtered-Reads- S_n ” give the number of input reads that were assigned to the n^{th} species after species-separation, where species are numbered in the order that they were specified on the command-line in Step 25. Dividing the sum of these numbers for all species except one by the number for that species, gives the approximate ratio of RNA of all the other species compared with that one; in the present case, dividing the sum of “Filtered-Reads” values for mouse and human by the value for rat gives the approximate ratio of “mouse plus human” to rat RNA. Then calculate a mean ratio of “mouse plus human” to rat RNA across all 3-species samples.

33 | For each 2-species [AU: Edit correct?] sample, calculate, for each rat gene, the percentage, p_i , of that gene’s expression likely to have arisen from mis-assigned rat reads:

$$p_i = 100 \times \frac{FPKM_{-rn_i} \times d}{FPKM_{-mm_{-}hs_i}},$$

where $FPKM_{-mm_{-}hs_i}$ and $FPKM_{-rn_i}$ are the approximate per-gene expression values calculated in Steps 30 and 31, and d is the mean ratio across samples of “mouse plus human” to rat RNA calculated in Step 32.

34 | Discard all rat genes for which the mean percentage of expression, calculated across a subset of 3-species samples, that is likely to have spuriously arisen from mouse or human RNA, is greater than some user-defined threshold, for example 10%, as was used previously ⁷. The number of genes included in this category is likely to be small.

▲ CRITICAL STEP Note that the exact subset of mixed-species samples used in this calculation must be chosen according to the differential expression comparison that is being made. In the example of microglial genes induced or repressed by LPS in the presence of neurons and astrocytes, the objective of the present step is to avoid reads actually originating from mouse and human genes, whose own expression may have been up-regulated or down-regulated as a result of being co-cultured with microglia in the presence of LPS, being mis-interpreted as indicating increases or decreases in the expression level of rat microglial genes. In the case of up-regulation, the mixed-species samples used to calculate the mean value of p_i should be those whereby LPS had been applied; whereas in the case of down-regulation, the control condition mixed-species samples should be used to calculate the mean value of p_i (note that in both cases, spurious assignment of human and mouse reads in the samples from the alternative condition might cause genes whose expression has truly changed as a consequence of the presence of LPS to not be recognized as having been differentially expressed, but it cannot lead to false differential expression calls).

TROUBLESHOOTING

Step	Problem	Potential explanation	Solution
Equipment setup	<i>wget</i> command cannot be found, preventing download of genome sequence and annotation files	<i>wget</i> is not installed	Install <i>wget</i> using a package manager appropriate to the operating system being used, for example <i>apt</i> on Linux distributions, or <i>Homebrew</i> on macOS. Alternatively, genome sequence and annotation files can be downloaded via FTP

Equipment setup	The <i>pip</i> command cannot be found	<i>pip</i> is not installed	<i>pip</i> should already be installed if version 2.7.9 or greater of Python 2 or version 3.4 or greater of Python 3 is being used. If not, follow the <i>pip</i> installation instructions found in https://pip.pypa.io/en/latest/installing/
10	Low yield of primary human astrocytes after thawing	The temperature change was too rapid when the cells were frozen or too slow when they were thawed	Use the Mr. Frosty container to freeze and thaw cells rapidly and safely; thawing should not take longer than 5 min
17	Uneven density of mouse neurons between the wells	Neurons were not mixed well enough before being plated	Continuously homogenize the cell suspension throughout the plating process
18	Mouse neurons are unhealthy or dying	The passage number of human astrocytes is too high, negatively affecting the vitality of mouse neurons	Thaw and recover human astrocytes characterized by a lower passage number
19	The medium turns pink after shaking, and a low yield of microglia is observed	<p>The caps of the flasks are not tight enough or the parafilm is not wrapped properly</p> <p>The flasks are not properly taped to the bottom of the shaker</p>	<p>Make sure caps are tight and parafilm blocks the filters of the caps to prevent the loss of CO₂</p> <p>Place the flasks on an adhesive matt to increase stability during shaking</p>
23	Low yield of total RNA	<p>The drug dose is too high or the inflammatory challenge is too protracted time-wise</p> <p>Cells are unhealthy for unknown reasons</p>	<p>Reduce the drug dose and/or shorten the stimulation time</p> <p>Verify the health of cells visually before applying the drug and repeat this task regularly through the protocol, including immediately prior to RNA extraction</p>
26	An error occurs during genome index generation in the <i>Sargasso</i> pipeline	<p>Insufficient disk space</p> <p>Incompatible input files</p>	<p>The <i>STAR</i> read aligner requires around 100Gb disk space to generate a genome index for a typical mammalian genome; ensure that enough disk space is available to generate indices for all the species used in the <i>Sargasso</i> pipeline</p> <p>The <i>STAR</i> read aligner requires that chromosome names in the supplied FASTA sequence files and the annotations GTF file match exactly; we thus recommend to use input files originating from the same source (for example both from Ensembl or both from UCSC).</p>

	An error occurs during read mapping in the <i>Sargasso</i> pipeline	Insufficient RAM available	<i>STAR</i> requires approximately 30Gb of RAM to align reads to mammalian genomes; execute the <i>Sargasso</i> pipeline on computing hardware of sufficient specifications
	<i>Sargasso</i> scripts cannot be found during pipeline execution	The Python virtual environment has not been activated	If a Python virtual environment is being used, activate the environment before executing the <i>Sargasso</i> pipeline
	Other error occurs during pipeline execution	Incorrect command issued or outdated versions of software being used	Check the <i>Sargasso</i> command issued, and that the versions of the <i>STAR</i> read aligner and <i>sambamba</i> alignment processing tool are up-to-date. If the error persists, post a bug report to the <i>Sargasso</i> issues list (https://github.com/statbio/Sargasso/issues)

•TIMING

Steps 1–7, Dissection and dissociation of rat and mouse cortices: 2–3 h

Step 8, Plating of rat cortical cells: 3 h

Step 9, Culture of rat microglia: 12–14 days

Step 10, Recovery of human astrocytes: 1–2 h

Steps 11–15, Passage, plate and culture human astrocytes: 3–4 days

Step 16, Culture of mouse neurons: 2–3 h

Steps 17–18, Co-culture of mouse neurons with human astrocytes: 7–8 days

Steps 19–20, Harvest and co-culture of microglia with mouse neurons and human astrocytes: 2–3 h

Step 21, Immunostaining: 2 days

Steps 22, Culture and inflammatory challenge: 4 days

Step 23, Harvest of RNA and RNA quality control: 3–4 h

Step 24, Preparation for *in silico* read separation: 30 min

Step 25, *In silico* read separation with *Sargasso*: 6.5 h

Steps 26–27, Examining the *Sargasso* output: 30 min

Steps 28–29, Downstream analysis: depends on the particular analyses that are to be performed

Steps 30–34, Avoiding misinterpretation of data: 2 h

Note that the timing information given here for the operation of *Sargasso* (Step 25) is approximate, as the execution time depends on the exact nature of the computing hardware available to the user, the number and identity of species being separated, the number of samples being processed, and the depth to which sequencing was performed. In the actual experiment the timing information provided was based on, 16 2.3 GHz processing cores were utilized, using a mixed mouse, rat, and human RNA-seq sample containing approximately 240 million 75 base pair paired-end reads; in this case, the execution time was 6.5 hours, at a rate of approximately 36 million reads processed per hour. Please note that, of this time, approximately 2/3 was taken up by

mapping reads to the three genomes with *STAR*. Increasing (or decreasing) the number of species to be separated will decrease (or increase) the rate at which samples are separated, as will decreasing or increasing the number of processing cores that are allocated.

ANTICIPATED RESULTS

We outline below the results of the analyses we performed in an example experiment designed to illustrate how it is possible to identify the transcriptional changes induced in neurons and in astrocytes by microglia, either unchallenged or activated by LPS (500 ng/ml) for 16 h), which triggers toll-like receptor 4 signaling. This follows a three species read separation (rat/mouse/neuron) that when employed in *Sargasso* leads to accurate FPKM and fold-induction values (Supplementary Results, Supplementary Figures 1-3). This application of the protocol enables microglia-dependent transcriptional changes to be tracked in both other cell types simultaneously. It also provides the means to determine the LPS-activated response of the microglia themselves. As noted above, the application differs from the previous published use of this protocol which involved the use of just two cell types ⁷. However, the inclusion of a third cell type, while modifying the length of the procedure, involves no conceptual changes to the workflow, nor does it impact on the results obtained. The experimental conditions and co-culture composition of the samples are described in Table 1.

TABLE 1. Co-culture composition and stimulation employed for the sample sets in Anticipated Results.

Sample	Neurons (mouse)	Astrocytes (human)	Microglia (rat)	LPS
1a	+	+	–	–
1b	+	+	–	+
2a	+	+	+	–
2b	+	+	+	+
3a	+	–	+	–
3b	+	–	+	+
4a	–	+	+	–
4b	–	+	+	+

Fig. 2a shows the overall proportion of reads within 3-species co-culture in sample set 2a and 2b attributable to the three different cell types. Using the rat (microglia) reads, it is possible to identify the transcriptional changes that microglia themselves undergo in response to LPS, by comparing rat reads within sample sets from 2a with those from 2b (see Table 1 for sample details). As expected, large numbers of genes were induced and repressed by LPS (Fig. 2b). By mapping the rat-microglia-free, astrocyte–neuron co-culture reads (LPS-treated, 1b samples (Table 1)) against

the rat genome, LPS-regulated genes with a >10% contribution from erroneous mapping from other species can be identified. In this example, only 14 out of the 1430 LPS-repressed genes (>1.5-fold repressed i.e. >33% reduction) and none of the 1075 induced genes (>1.5-fold increased) were identified as erroneously mapped from other species, and discarded from the final scatter plot (Fig. 2b). Induced genes included well-known LPS/TLR4-inducible cytokine/chemokine genes *Cxcl1*, *Cxcl2*, *Cxcl3*, *Ccl3*, *Il1a* and *Il1b* (Fig. 2b). Among the genes repressed was the anti-inflammatory marker *Igf1*. Using co-cultures of microglia, neurons and astrocytes from single species [AU: Edit correct?], we confirmed as far as possible that the LPS-induced microglial genes observed to be up- and down-regulated in the presence of neurons and astrocytes were not due to the neurons and astrocytes being from a different species than the microglia (Supplementary Results, Supplementary Figure 5).

In order to determine the non-cell-autonomous influence of unchallenged microglia on the astrocytic transcriptome, one can compare human reads within sample sets from 2a with those from 1a. As is illustrated (Fig. 2c) very few astrocytic genes are altered (only 10). Nevertheless, one should still check if anyone of them could be due to erroneous mapping, by mapping the human astrocyte-free neuron–microglia co-culture reads (sample set 3a) against the human genome (in this instance none of the 10 genes had a contribution from erroneous mapping from other species). Since LPS-challenged microglia release factors that induce phenotypic and transcriptional changes in astrocytes¹⁸, we expected that the influence of LPS-challenged microglia on the astrocytic transcriptomes would be profound. To ascertain this, human reads within sample sets from 2b were compared with those from 1b. Note that it is important to compare samples with 1b, rather than 1a, in order to control for the direct influence that LPS has on astrocytes, and distinguish this influence from that of microglia. As expected, a large number of genes were induced and repressed. Any genes arising due to potential erroneous read mapping can be identified by mapping the human-astrocyte-free neuron–microglia co-culture reads (LPS-treated, 3b samples) against the human genome. In this instance none of the (1039) genes reported as induced (>1.5 fold increase in expression), and three of the 477 genes repressed (>1.5 fold decrease in expression) had a >10% contribution from erroneous mapping from other species, and were as a consequence discarded from our scatter graph (Fig. 2d). Genes induced in astrocytes by LPS-activated microglia include markers of neurotoxic “A1” type astrocytes *C3* and *MX1*¹⁸ (Fig. 2d), as well as *CHI3L1* and *ICAM1*, which are known to be aberrantly expressed by astrocytes in a range of neurodegenerative diseases^{48,49}.

To determine the influence of unchallenged microglia on the neuronal transcriptome, reads from the 2a sample sets can be compared with those from 1a, but this time focusing on mouse reads (Fig. 2e). No neuronal genes were found to be altered by unchallenged microglia (fold-change cut-off of 1.5-fold and an adjusted p-value (false discovery rate) of 0.05). To determine the influence that LPS-challenged microglia had on the neuronal transcriptome, mouse reads from 2b sample sets can be compared with those from 1b. As expected, this comparison revealed a large

number of induced genes (Fig. 2f). Controlling for erroneous mapping is achieved by mapping the neuron-free astrocyte/microglia co-culture reads (LPS-treated, 4b samples) against the mouse genome. This confirms very limited erroneous mapping: only two out of 430 induced genes and zero of the 9 repressed genes had a >10% contribution from erroneous mapping from other species, and so were discarded from our lists. Genes induced in neurons by LPS-activated microglia include known interferon-responsive genes *Oasl2*, *Isg15*, *Ifitm3*, *Timp1*, and *Stat2* (Fig. 2f).

Thus, in the example application of the protocol just described, our approach revealed simultaneously the microglial response to LPS and the consequent influence of this microglial response on astrocytic and neuronal transcriptomes. Of note, LPS-activated microglia induce quite distinct transcriptional programs in neurons and astrocytes. If one takes the genes induced by microglia in astrocytes (Fig. 2d) and neurons (Fig. 2f) and plots the fold change of those genes in the other cell type, it is clear that a relatively small number of genes are induced >1.5-fold in both cell types (Supplementary Figure 6). To conclude, the mixed species co-culture, followed by RNA-seq and *Sargasso* read separation can be implemented together to probe transcriptional profiles of neurons, astrocytes, and microglia within the same sample set, an approach that may be useful in answering questions surrounding cell-autonomous and non-autonomous signaling during development and in healthy and diseased states.

Author contributions statements

JQ, PB, and PH developed and validated the mixed-species co-culture systems. OD, SH, and TIS developed the *Sargasso* tool and analyzed the data. GEH conceived and directed the project, and analyzed the data. GEH, JQ, and OD wrote the manuscript.

Acknowledgments

This work is funded by UK Dementia Research Institute partner funders the Medical Research Council, Alzheimer's Research UK, and Alzheimer's Society, as well as the Wellcome Trust, and the Simons Initiative for the Developing Brain. We are grateful to Xin He for comments on the manuscript.

Competing Interest Statement

The authors declare that they have no competing financial or non-financial interests relating to this manuscript.

References

- 1 Denef, C. Contact-dependent Signaling. *Cell Communication Insights* **6**, 1-11, doi:10.4137/CCi.s13924. (2014).
- 2 Bell, K. F. & Hardingham, G. E. The influence of synaptic activity on neuronal health. *Curr Opin Neurobiol* **21**, 299-305, doi:S0959-4388(11)00006-7 [pii] 10.1016/j.conb.2011.01.002 (2011).

- 3 West, A. E. & Greenberg, M. E. Neuronal activity-regulated gene transcription in synapse development and cognitive function. *Cold Spring Harb Perspect Biol* **3**, doi:cshperspect.a005744 [pii] 10.1101/cshperspect.a005744 (2011).
- 4 Hardingham, G. E. & Lipton, S. A. Regulation of neuronal oxidative and nitrosative stress by endogenous protective pathways and disease processes. *Antioxid Redox Signal* **14**, 1421-1424, doi:10.1089/ars.2010.3573 (2011).
- 5 Bell, K. F. & Hardingham, G. E. CNS peroxiredoxins and their regulation in health and disease. *Antioxid Redox Signal* **14**, 1467-1477, doi:10.1089/ars.2010.3567 (2011).
- 6 Baxter, P. S. *et al.* Synaptic NMDA receptor activity is coupled to the transcriptional control of the glutathione system. *Nat Commun* **6**, 6761, doi:ncomms7761 [pii] 10.1038/ncomms7761 (2015).
- 7 Hasel, P. *et al.* Neurons and neuronal activity control gene expression in astrocytes to regulate their development and metabolism. *Nat Commun* **8**, 15132, doi:10.1038/ncomms15132 (2017).
- 8 Mensch, S. *et al.* Synaptic vesicle release regulates myelin sheath number of individual oligodendrocytes in vivo. *Nat Neurosci* **18**, 628-630, doi:10.1038/nn.3991 (2015).
- 9 Hardingham, G. E. & Do, K. Q. Linking early-life NMDAR hypofunction and oxidative stress in schizophrenia pathogenesis. *Nat Rev Neurosci* **17**, 125-134, doi:10.1038/nrn.2015.19 (2016).
- 10 Bell, K. F., Fowler, J. H., Al-Mubarak, B., Horsburgh, K. & Hardingham, G. E. Activation of Nrf2-regulated glutathione pathway genes by ischemic preconditioning. *Oxid Med Cell Longev* **2011**, 689524, doi:10.1155/2011/689524 (2011).
- 11 Baxter, P. S. & Hardingham, G. E. Adaptive regulation of the brain's antioxidant defences by neurons and astrocytes. *Free Radic Biol Med* **100**, 147-152, doi:10.1016/j.freeradbiomed.2016.06.027 (2016).
- 12 Allen, N. J. *et al.* Astrocyte glypicans 4 and 6 promote formation of excitatory synapses via GluA1 AMPA receptors. *Nature* **486**, 410-414, doi:nature11059 [pii] 10.1038/nature11059 (2012).
- 13 Clarke, L. E. & Barres, B. A. Emerging roles of astrocytes in neural circuit development. *Nat Rev Neurosci* **14**, 311-321, doi:10.1038/nrn3484 (2013).
- 14 Schafer, D. P. & Stevens, B. Microglia Function in Central Nervous System Development and Plasticity. *Cold Spring Harb Perspect Biol* **7**, a020545, doi:10.1101/cshperspect.a020545 (2015).
- 15 Salter, M. W. & Stevens, B. Microglia emerge as central players in brain disease. *Nat Med* **23**, 1018-1027, doi:10.1038/nm.4397 (2017).
- 16 Perry, V. H. & Holmes, C. Microglial priming in neurodegenerative disease. *Nat Rev Neurol* **10**, 217-224, doi:nrneurol.2014.38 [pii] 10.1038/nrneurol.2014.38 (2014).
- 17 Hoarau, J. J. *et al.* Activation and control of CNS innate immune responses in health and diseases: a balancing act finely tuned by neuroimmune regulators (NIReg). *CNS Neurol Disord Drug Targets* **10**, 25-43 (2011).
- 18 Liddel, S. A. *et al.* Neurotoxic reactive astrocytes are induced by activated microglia. *Nature*, doi:10.1038/nature21029 nature21029 [pii] (2017).
- 19 Miron, V. E. *et al.* M2 microglia and macrophages drive oligodendrocyte differentiation during CNS remyelination. *Nat Neurosci* **16**, 1211-1218, doi:nn.3469 [pii] 10.1038/nn.3469 (2013).
- 20 Baxter, P. S., Martel, M. A., McMahon, A., Kind, P. C. & Hardingham, G. E. Pituitary adenylate cyclase-activating peptide induces long-lasting neuroprotection through the induction of activity-dependent signaling via the cyclic AMP response element-binding protein-regulated transcription co-activator 1. *J Neurochem* **118**, 365-378, doi:10.1111/j.1471-4159.2011.07330.x (2011).
- 21 Bell, K. F. *et al.* Neuronal development is promoted by weakened intrinsic antioxidant defences due to epigenetic repression of Nrf2. *Nat Commun* **6**, 7066, doi:ncomms8066 [pii] 10.1038/ncomms8066 (2015).

- 22 Edman, S. *et al.* TCN 201 selectively blocks GluN2A-containing NMDARs in a GluN1 co-agonist dependent but non-competitive manner. *Neuropharmacology* **63**, 441-449, doi:S0028-3908(12)00163-3 [pii]
10.1016/j.neuropharm.2012.04.027 (2012).
- 23 Wilhelm, B. T., Marguerat, S., Goodhead, I. & Bahler, J. Defining transcribed regions using RNA-seq. *Nat Protoc* **5**, 255-266, doi:10.1038/nprot.2009.229 (2010).
- 24 Okaty, B. W., Sugino, K. & Nelson, S. B. A quantitative comparison of cell-type-specific microarray gene expression profiling methods in the mouse brain. *PLoS ONE* **6**, e16493, doi:10.1371/journal.pone.0016493 (2011).
- 25 Okaty, B. W., Sugino, K. & Nelson, S. B. Cell type-specific transcriptomics in the brain. *J Neurosci* **31**, 6939-6943, doi:31/19/6939 [pii]
10.1523/JNEUROSCI.0626-11.2011 (2011).
- 26 Doyle, J. P. *et al.* Application of a translational profiling approach for the comparative analysis of CNS cell types. *Cell* **135**, 749-762, doi:S0092-8674(08)01366-4 [pii]
10.1016/j.cell.2008.10.029 (2008).
- 27 Ahdesmaki, M. J., Gray, S. R., Johnson, J. H. & Lai, Z. Disambiguate: An open-source application for disambiguating two species in next generation sequencing data from grafted samples. *F1000Res* **5**, 2741, doi:10.12688/f1000research.10082.2 (2016).
- 28 Conway, T. *et al.* Xenome--a tool for classifying reads from xenograft samples. *Bioinformatics* **28**, i172-178, doi:10.1093/bioinformatics/bts236 (2012).
- 29 Kumar, S., Stecher, G., Suleski, M. & Hedges, S. B. TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. *Molecular biology and evolution* **34**, 1812-1819, doi:10.1093/molbev/msx116 (2017).
- 30 Puddifoot, C. *et al.* PGC-1alpha negatively regulates extrasynaptic NMDAR activity and excitotoxicity. *J Neurosci* **32**, 6995-7000, doi:32/20/6995 [pii]
10.1523/JNEUROSCI.6407-11.2012 (2012).
- 31 Chow, J. C., Young, D. W., Golenbock, D. T., Christ, W. J. & Gusovsky, F. Toll-like receptor-4 mediates lipopolysaccharide-induced signal transduction. *J Biol Chem* **274**, 10689-10692 (1999).
- 32 Qiu, J. *et al.* Evidence for evolutionary divergence of activity-dependent gene expression in developing neurons. *Elife* **5**, doi:10.7554/eLife.20337
e20337 [pii] (2016).
- 33 Espuny-Camacho, I. *et al.* Hallmarks of Alzheimer's Disease in Stem-Cell-Derived Human Neurons Transplanted into Mouse Brain. *Neuron* **93**, 1066-1081 e1068, doi:10.1016/j.neuron.2017.02.001 (2017).
- 34 Wang, S. *et al.* Human iPSC-derived oligodendrocyte progenitor cells can myelinate and rescue a mouse model of congenital hypomyelination. *Cell Stem Cell* **12**, 252-264, doi:10.1016/j.stem.2012.12.002 (2013).
- 35 Han, X. *et al.* Forebrain engraftment by human glial progenitor cells enhances synaptic plasticity and learning in adult mice. *Cell Stem Cell* **12**, 342-353, doi:10.1016/j.stem.2012.12.015 (2013).
- 36 O'Doherty, A. *et al.* An aneuploid mouse strain carrying human chromosome 21 with Down syndrome phenotypes. *Science* **309**, 2033-2037, doi:309/5743/2033 [pii]
10.1126/science.1114535 (2005).
- 37 Kim, K. M., Abdelmohsen, K., Mustapic, M., Kapogiannis, D. & Gorospe, M. RNA in extracellular vesicles. *Wiley Interdiscip Rev RNA* **8**, doi:10.1002/wrna.1413 (2017).
- 38 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:bts635 [pii]
10.1093/bioinformatics/bts635 (2013).
- 39 Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J. & Prins, P. Sambamba: fast processing of NGS alignment formats. *Bioinformatics* **31**, 2032-2034, doi:10.1093/bioinformatics/btv098 (2015).
- 40 Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* **14**, 417-419, doi:10.1038/nmeth.4197 (2017).

- 41 Soneson, C., Love, M. I. & Robinson, M. D. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res* **4**, 1521, doi:10.12688/f1000research.7563.2 (2015).
- 42 Yates, A. *et al.* Ensembl 2016. *Nucleic Acids Res* **44**, D710-716, doi:10.1093/nar/gkv1157 (2016).
- 43 McKenzie, G. J. *et al.* Nuclear Ca²⁺ and CaM kinase IV specify hormonal- and Notch-responsiveness. *J Neurochem* **93**, 171-185, doi:JNC3010 [pii] 10.1111/j.1471-4159.2005.03010.x (2005).
- 44 Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930, doi:10.1093/bioinformatics/btt656 (2014).
- 45 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).
- 46 Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140, doi:10.1093/bioinformatics/btp616 (2010).
- 47 Anders, S. *et al.* Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. *Nat Protoc* **8**, 1765-1786, doi:10.1038/nprot.2013.099 (2013).
- 48 Akiyama, H. *et al.* Expression of intercellular adhesion molecule (ICAM)-1 by a subset of astrocytes in Alzheimer disease and some other degenerative neurological disorders. *Acta Neuropathol* **85**, 628-634 (1993).
- 49 Bonneh-Barkay, D., Wang, G., Starkey, A., Hamilton, R. L. & Wiley, C. A. In vivo CHI3L1 (YKL-40) expression in astrocytes in acute and chronic neurological diseases. *J Neuroinflammation* **7**, 34, doi:10.1186/1742-2094-7-34 (2010).
- 50 Muffat, J. *et al.* Efficient derivation of microglia-like cells from human pluripotent stem cells. *Nat Med* **22**, 1358-1367, doi:10.1038/nm.4189 (2016).

Figure Legends

Figure 1. Determining the presence of expected cell markers and the absence of markers of contaminating cell types. Immunofluorescent staining of the indicated cell types, alone or in combination, in mono- or three-way co-culture, as indicated. In all cases DAPI-stained nuclei are pseudo-colored red, and the neuronal (Neuro-Chrom), astrocytic (Gfap) and microglial (Iba1) marker in green. Scale bar: 50 μ m.

Figure 2. Differential gene expression in microglia, astrocytes, and neurons in co-culture. For both the three-species co-culture (mouse neurons, human astrocytes, and rat microglia) and the mouse neuron–human astrocyte two-species co-culture, the number of reads unambiguously mapped to each species is shown, expressed as a percentage of the total number of unambiguously mapped reads of all species (n=3 biological replicates, defined as utilizing primary tissue from distinct animals here and throughout the manuscript). For the two-species co-culture samples 64.1 \pm 2.5 million reads per sample were unambiguously mapped, and for the three-species co-culture samples 133.7 \pm 12.5 million reads per sample were unambiguously mapped. Error bars represent the standard error of the mean. **B)** LPS-induced microglial gene expression. Species-specific read sorting identified rat (i.e. microglial) reads. Expression of genes (FPKM) in microglia \pm LPS is plotted for the genes expressed >0.5 FPKM averaged over the conditions. Red crosses indicate the microglial genes whose expression level was modified by LPS presence >1.5 -fold (*DESeq2* $P_{adj} < 0.05$, n=3 biological replicates, 1075 genes induced, 1416 genes repressed (14 of the initial 1430 genes called as being significantly changed were discarded due to erroneous mapping)) Grey data points indicate the genes whose expression level modification falls below these thresholds. Circled and named genes are referred to in the text. **C,D)** Microglia-dependent transcriptional changes induced in astrocytes. Either three-species or two-species (lacking

microglia) co-cultures were treated \pm LPS and RNA-seq performed. Species-specific read sorting identified human (i.e. astrocyte) reads. Expression of genes (FPKM) in astrocytes in the presence of neurons only (two-species co-culture, x-axis) is plotted against gene expression in astrocytes when microglia were also present (three-species co-culture, y-axis). Genes significantly altered in their level of expression by the presence of microglia are marked by red crosses (*DESeq2* $P_{\text{adj}} < 0.05$, $n=3$). (C) refers to basal conditions and (D) to LPS-treated cultures (1039 genes induced, 474 genes repressed (3 of the initial 477 genes called as being significantly changed were discarded due to erroneous mapping)). Circled and named genes are referred to in the text. **E,F)** Microglia-dependent transcriptional changes induced in neurons. The same samples were used as in C,D, and species-specific read sorting employed to identify mouse (neuronal) reads. Genes significantly altered by the presence of microglia are marked by red crosses (*DESeq2* $P_{\text{adj}} < 0.05$, $n=3$). (E) refers to basal conditions and (F) to LPS-treated cultures (430 genes induced, 9 genes repressed). Circled and named genes are referred to in the text.

Supplementary Information:

Supplementary Results.

Supplementary Figure 1. Impact of *Sargasso* species disambiguation on gene expression quantification. **A)** Single-species RNA-seq reads (rat samples MGLmonoCTR1–3 from ArrayExpress accession E-MTAB-5987 (this study), human samples GSM2285374–7 from Gene Expression Omnibus series GSE85839⁵⁰ and mouse samples CTR1-34316426, CTR2-34335325 and CTR3-34312414 from ArrayExpress accession E-MTAB-5489³²), were taken, and we calculated the % reads lost for each protein-coding gene expressed >1 FPKM (by normal mapping—defined as mapping which requires perfect match to target species, but no disambiguation between any other species) when performing the *Sargasso* pipeline requiring disambiguation of reads from the other two species. A cumulative distribution plot for genes against % of reads lost is shown. **B,C)** For rat RNA-seq reads (rat samples MGLmonoCTR1–3 from E-MTAB-5987), gene FPKM values were calculated using normal mapping and *Sargasso* mapping, requiring disambiguation of reads from mouse and human genomes. FPKM values were plotted against each other for all 12,432 protein-coding genes expressed >1 FPKM (by normal mapping) (B). In (C), the FPKM (*Sargasso*) was calculated as a % of the FPKM (normal mapping), and a frequency distribution of the genes shown (with 5% bin size). **D,E)** For human RNA-seq reads (samples GSM2285374–7 from GSE85839), gene FPKM values were calculated using normal mapping and *Sargasso* mapping, requiring disambiguation of reads from rat and mouse genomes. FPKM values were plotted against each other for all 11,668 protein-coding genes expressed >1 FPKM (by normal mapping) (D). In (E), the FPKM (*Sargasso*) was calculated as a % of the FPKM (normal mapping), and a frequency distribution of the genes shown (with 5% bin size). **F,G)** For mouse RNA-seq reads (mouse samples CTR1-34316426, CTR2-34335325 and CTR3-34312414 from E-MTAB-5489), gene FPKM values were calculated using normal mapping and *Sargasso* mapping, requiring disambiguation of reads from rat and human genomes. FPKM values were plotted against each other for all 11,971 protein-coding genes expressed >1 FPKM (by normal mapping) (F). In (G), the FPKM (*Sargasso*) was calculated as a % of the FPKM (normal mapping), and a frequency distribution of the genes shown (with 5% bin size).

Supplementary Figure 2. Impact of the *Sargasso* species disambiguation on transcript expression quantification. **A,B)** For rat RNA-seq reads (rat samples MGLmonoCTR1–3 from E-MTAB-5987), FPKM values of all transcripts of protein-coding genes (as annotated by Ensembl) were calculated using normal mapping and *Sargasso* mapping, requiring disambiguation of reads from mouse and human genomes. FPKM values were plotted against each other for all those 16,169 transcripts expressed >1 FPKM (by normal mapping) (A). In (B), the FPKM (*Sargasso*) was calculated as a % of the FPKM (normal mapping), and a frequency distribution of the transcripts shown. **C,D)** For human RNA-seq reads (samples GSM2285374–7 from GSE85839), FPKM values of all transcripts of protein-coding genes (as annotated by Ensembl) were calculated using normal mapping and *Sargasso* mapping, requiring disambiguation of reads from mouse and rat

genomes. FPKM values were plotted against each other for all those 34,149 transcripts expressed >1 FPKM (by normal mapping) (C). In (D), the FPKM (*Sargasso*) was calculated as a % of the FPKM (normal mapping), and a frequency distribution of the transcripts shown. **E,F**) For mouse RNA-seq reads (mouse samples CTR1-34316426, CTR2-34335325 and CTR3-34312414 from E-MTAB-5489), FPKM values of all transcripts of protein-coding genes (as annotated by Ensembl) were calculated using normal mapping and *Sargasso* mapping, requiring disambiguation of reads from human and rat genomes. FPKM values were plotted against each other for all those 34,024 transcripts expressed >1 FPKM (by normal mapping) (E). In (F), the FPKM (*Sargasso*) was calculated as a % of the FPKM (normal mapping), and a frequency distribution of the transcripts shown.

Supplementary Figure 3. Impact of *Sargasso* species disambiguation on stimulus-induced gene- and transcript- level fold induction quantification. **A,B**) For rat RNA-seq reads (rat control samples MGLmonoCTR1–3, and LPS-treated samples MGLmonoLPS1–3, E-MTAB-5987 (this publication-sample sets 6a and 6b: see *Supplementary Results*), DESeq2 Log₂-fold change (LPS vs. Con) was calculated both at the gene level (A) and transcript level (B), using normal mapping, and for *Sargasso* mapping requiring disambiguation of reads from mouse and human genomes. For all genes/transcripts expressed > 1FPKM on average across the samples (by normal mapping), DESeq2 Log₂-fold change for the two approaches was plotted against each other, and a correlation coefficient calculated. **C,D**) For mouse RNA-seq reads (mouse DIV4 cortical neuron control samples, and high K⁺-treated samples, both from E-MTAB-5489³²), DESeq2 Log₂-fold change (high K⁺ vs. Con) was calculated both at the gene level (C) and transcript level (D), using normal mapping, and for *Sargasso* mapping requiring disambiguation of reads from rat and human genomes. For all genes/transcripts expressed > 1FPKM on average across the samples (by normal mapping), DESeq2 Log₂-fold change for the two approaches was plotted against each other, and a correlation coefficient calculated. **E,F**) For human RNA-seq reads (human ES cell-derived neuron control samples, and high K⁺-treated samples, both from E-MTAB-5489³²), DESeq2 Log₂-fold change (high K⁺ vs. Con) was calculated both at the gene level (E) and transcript level (F), using normal mapping, and for *Sargasso* mapping requiring disambiguation of reads from rat and mouse genomes. For all genes/transcripts expressed > 1FPKM on average across the samples (by normal mapping), DESeq2 Log₂-fold change for the two approaches was plotted against each other, and a correlation coefficient calculated.

Supplementary Figure 4. *Sargasso* species disambiguation of human RNA-seq reads against species of varying evolutionary distance. **A**) Human RNA-seq reads (Gene Expression Omnibus samples GSM2285374–7⁵⁰), were taken, and we calculated the % reads lost for each protein-coding gene expressed >1 FPKM (by normal mapping – defined as mapping which requires perfect match to target species, but no disambiguation between any other species, 11,661 genes in total) when performing the *Sargasso* pipeline requiring disambiguation of reads against mouse, rat, macaque, or chimpanzee. A cumulative distribution plot for genes against % of reads lost is shown for each of the four *Sargasso* species disambiguation procedures. **B-E**) For the human RNA-seq reads in (A), FPKM values of all protein-coding genes were calculated using normal mapping and *Sargasso* mapping, requiring disambiguation of reads from mouse (B), rat (C), macaque (D) and chimpanzee (E). FPKM values were plotted against each other for all those 11,661 genes expressed >1 FPKM (by normal mapping), and the correlation coefficient calculated.

Supplementary Figure 5. Expression changes induced by LPS in microglia are not influenced by having other cells present from different species. **A,B**) Differential gene expression analysis between samples 5a and 5b, co-cultures of rat microglia, neurons and astrocytes (± LPS). We took the list of genes significantly induced (304 genes) or repressed (113 genes) >4-fold by LPS treatment in rat microglia in the mixed species co-culture (samples 2a vs. 2b) and then looked at the LPS-dependent regulation of the subset of these genes whose induction could be tracked in a single species co-culture (5a vs. 5b) by virtue of their expression being >5-fold higher in a pure microglial culture, than the mixed microglia-astrocyte-neuron co-culture (samples 5a vs. 6a), and expressed at least 1 FPKM in mono-cultured microglia. Applying these criteria meant that we could, to a first approximation, monitor the regulation of 108/304 LPS-

induced genes (A), and 44/113 LPS-repressed genes (B), in microglia in a single species microglia-neuron-astrocyte co-culture. The DESeq2 Log₂fold-change is shown for each of the 108 LPS-induced genes (A) and 44 LPS-repressed genes (B), and a *P* value calculated (paired t-test, samples 5a vs. 5b, n=108 (A), n=44 (B)).

Supplementary Figure 6. Activated microglia induce largely distinct specific transcriptional responses in neurons and astrocytes. For genes induced (>1.5-fold) by microglia in astrocytes (Fig. 2d) and in neurons (Fig. 2f), the fold change in neurons is plotted against that in astrocytes. The top-right quadrant formed by the crossed dotted lines includes those genes induced >1.5-fold by microglia in both neurons and astrocytes, a relatively small number of genes.

Supplementary Software 1. Sargasso-1.1.zip – contains version 1.1 of the Sargasso software and documentation (doi:10.5281/zenodo.260123).

Fig. 1

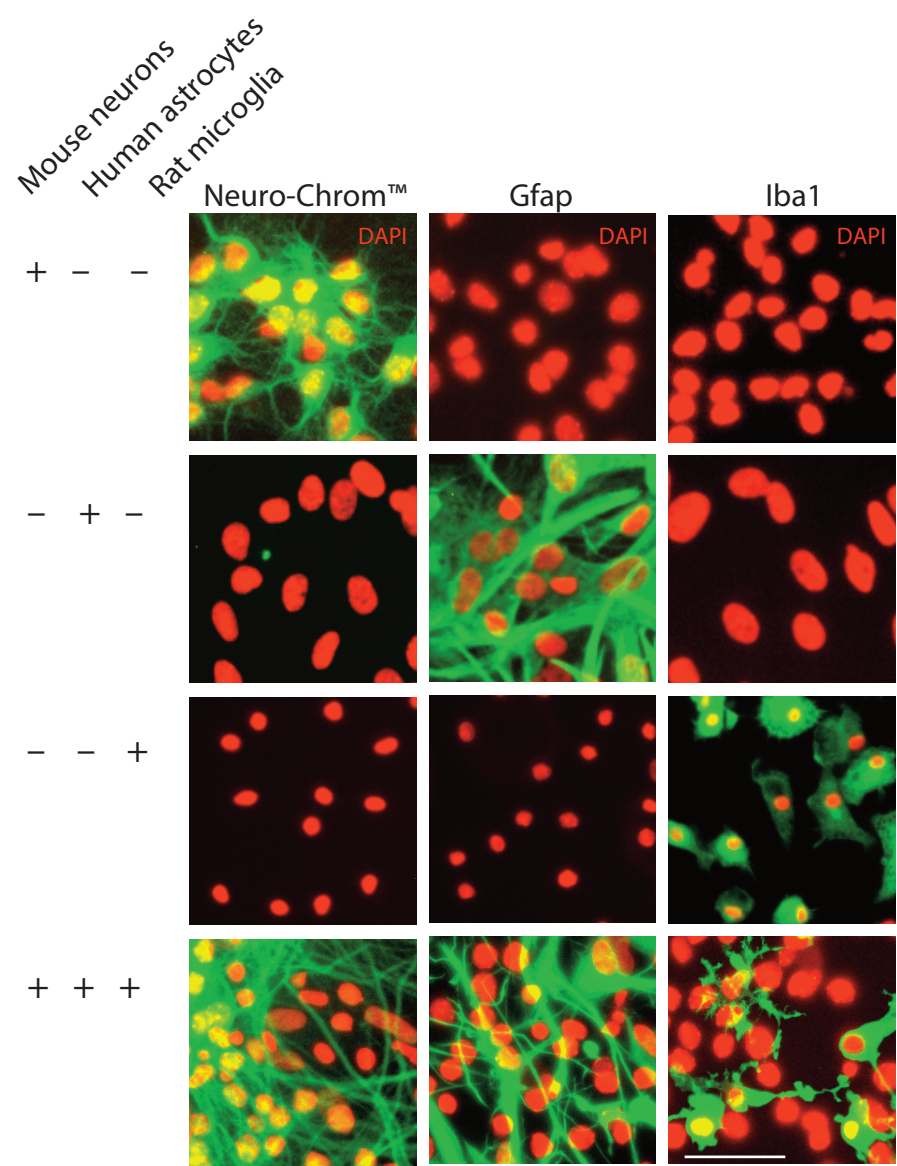


Fig. 2

