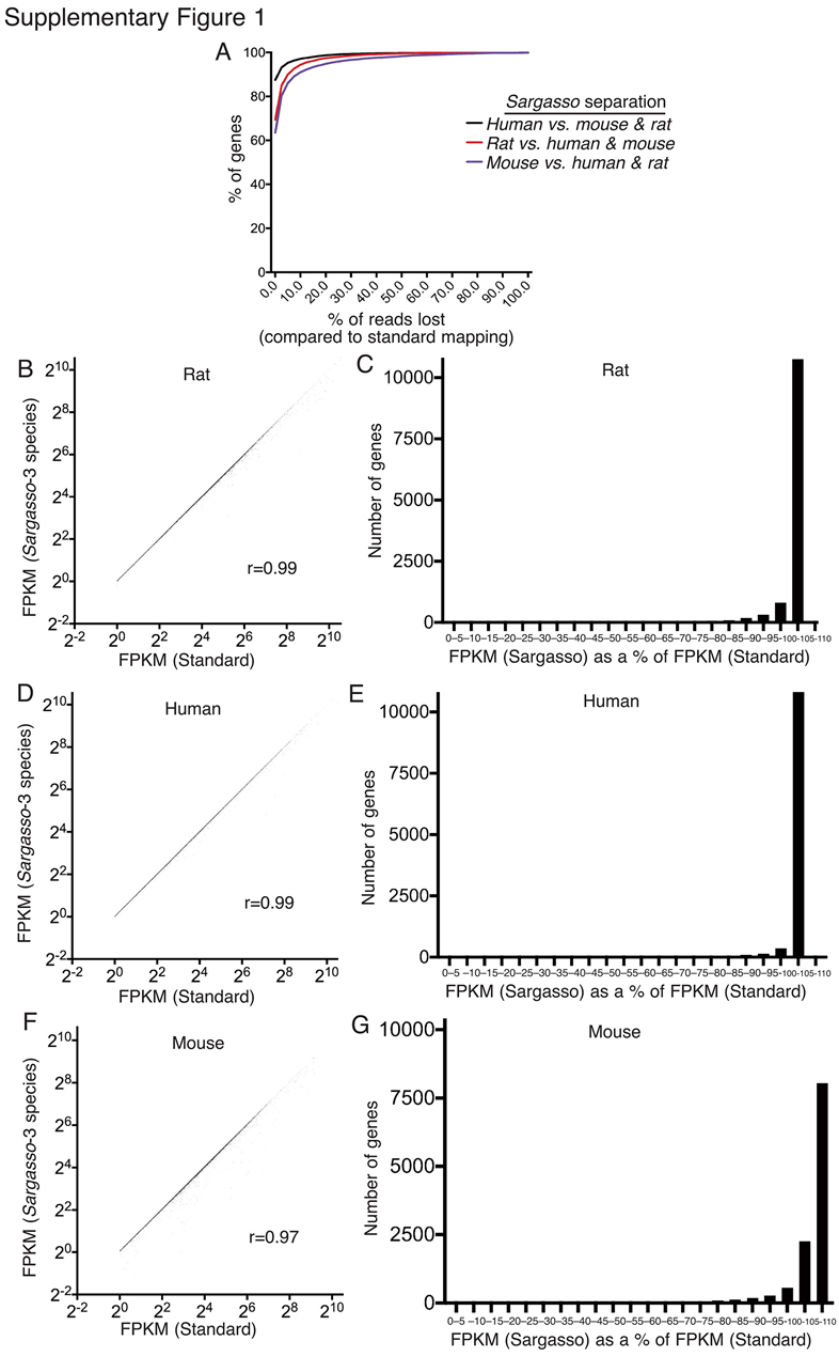
Mixed-species RNA-seq for elucidating non-cell-autonomous control of gene transcription

Supplemental Figures

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

Jing Qiu, Owen Dando, Paul Baxter, Philip Hasel, Samuel Heron, T. Ian Simpson, and Giles E. Hardingham

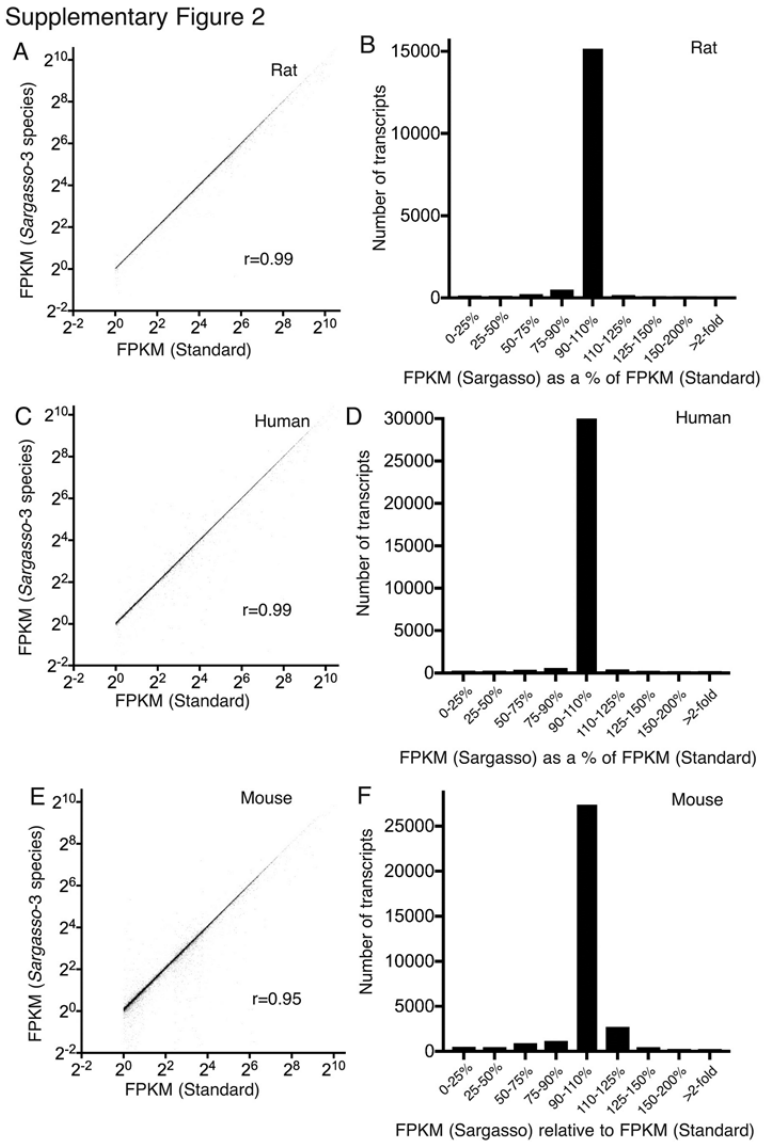


Supplementary Figure 1

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 GSE85839 [1] and mouse samples CTR1-34316426, CTR2-34335325 and CTR3-34312414 from ArrayExpress accession E-MTAB-5489 [2]), 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).



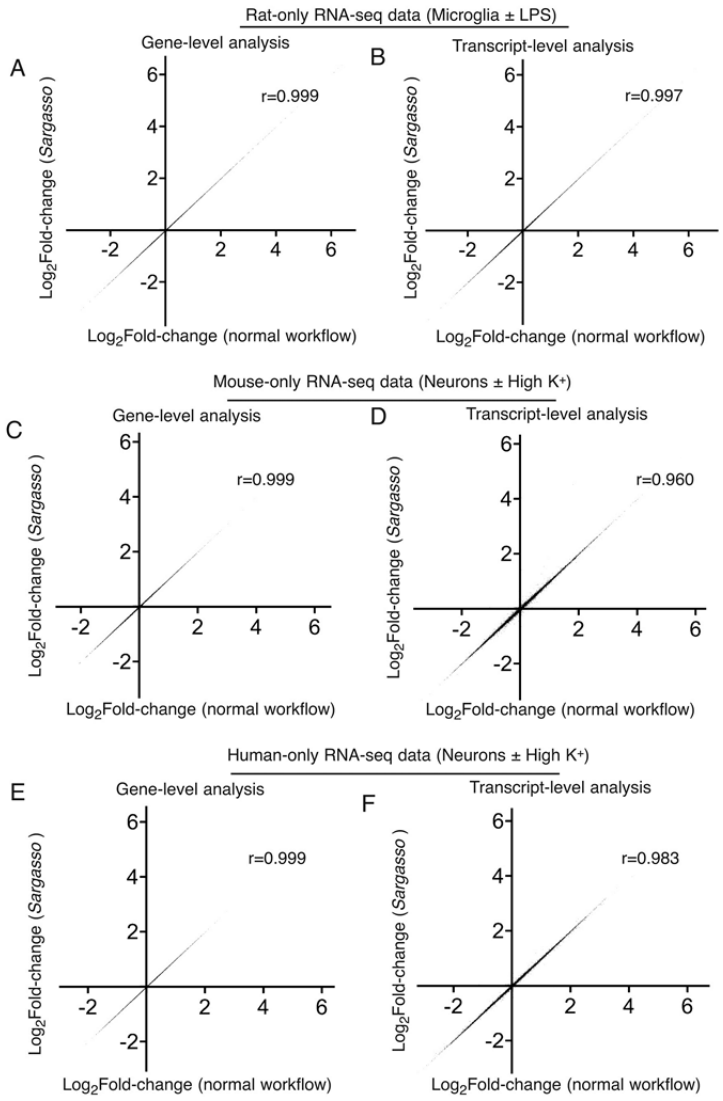
Supplementary Figure 2

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.

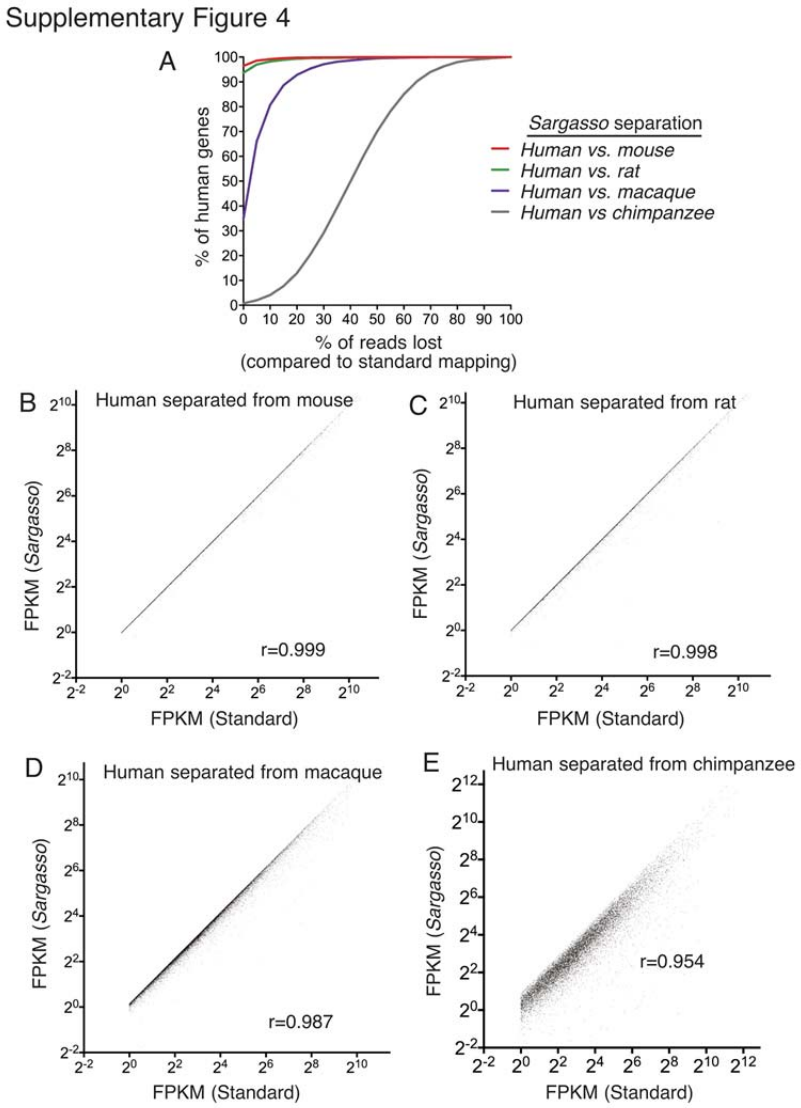
Supplementary Figure 3



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 Log_2-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 Log_2-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-5489 [2]), DESeq2 Log_2-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 Log_2-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-5489 [2]), DESeq2 Log_2-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 Log_2-fold change for the two approaches was plotted against each other, and a correlation coefficient calculated.

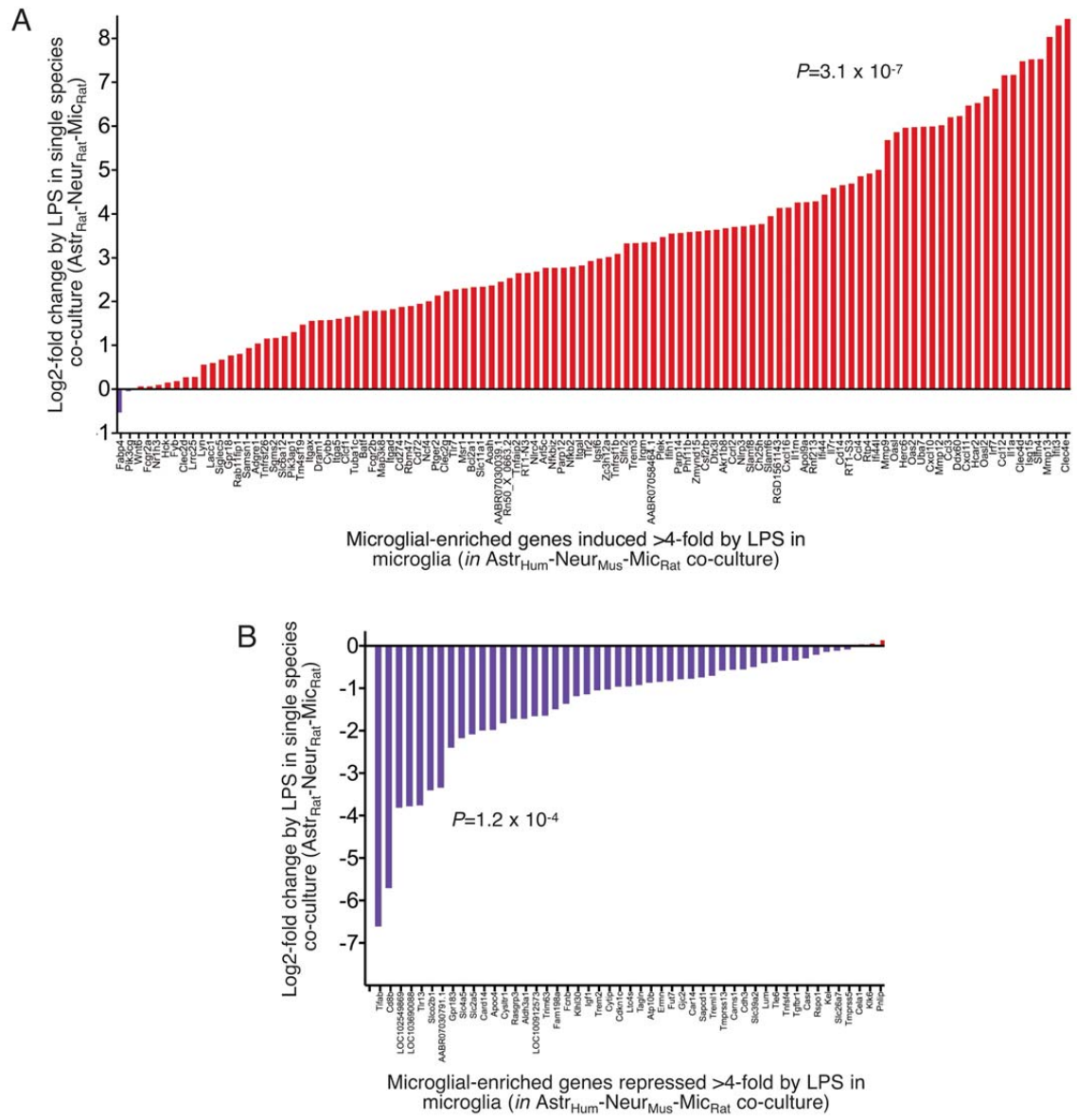


Supplementary Figure 4

***Sargasso* species disambiguation of human RNA-seq reads against species of varying evolutionary distance**

A) Human RNA-seq reads (Gene Expression Omnibus samples GSM2285374–7 [1]), 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.

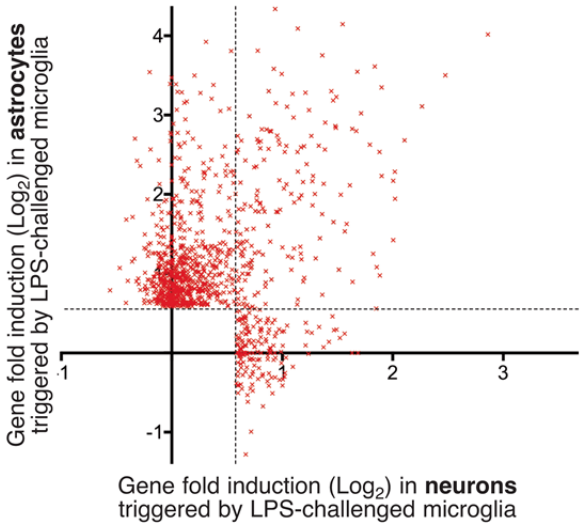
Supplementary Figure 5



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 Log_2fold-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.

References

1 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).

2 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).