

ARTICLE

Received 20 Aug 2015 | Accepted 14 Jan 2016 | Published 19 Feb 2016

DOI: 10.1038/ncomms10717

OPEN

# Gene expression in human brain implicates sexually dimorphic pathways in autism spectrum disorders

Donna M. Werling<sup>1</sup>, Neelroop N. Parikshak<sup>1</sup> & Daniel H. Geschwind<sup>1,2,3,4</sup>

Autism spectrum disorder (ASD) is more prevalent in males, and the mechanisms behind this sex-differential risk are not fully understood. Two competing, but not mutually exclusive, hypotheses are that ASD risk genes are sex-differentially regulated, or alternatively, that they interact with characteristic sexually dimorphic pathways. Here we characterized sexually dimorphic gene expression in multiple data sets from neurotypical adult and prenatal human neocortical tissue, and evaluated ASD risk genes for evidence of sex-biased expression. We find no evidence for systematic sex-differential expression of ASD risk genes. Instead, we observe that genes expressed at higher levels in males are significantly enriched for genes upregulated in post-mortem autistic brain, including astrocyte and microglia markers. This suggests that it is not sex-differential regulation of ASD risk genes, but rather naturally occurring sexually dimorphic processes, potentially including neuron-glia interactions, that modulate the impact of risk variants and contribute to the sex-skewed prevalence of ASD.

<sup>1</sup>Center for Neurobehavioral Genetics, Semel Institute, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, California 90095, USA. <sup>2</sup>Program in Neurogenetics, Department of Neurology, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, California 90095, USA. <sup>3</sup>Center for Autism Research and Treatment, Semel Institute, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, California 90095, USA. <sup>4</sup>Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, California 90095, USA. Correspondence and requests for materials should be addressed to D.H.G. (email: dhg@mednet.ucla.edu).

Autism spectrum disorder (ASD) is a developmental condition characterized by deficits in social communication and restricted, repetitive behaviours or interests<sup>1</sup> that is estimated to affect 1 in 68 children in the United States<sup>2</sup>. Genetic variation contributes strongly to ASD risk, as evidenced by high familial recurrence<sup>3–5</sup>, overlap with monogenic syndromes such as Fragile X (*FMRI*) or Smith–Lemli–Opitz syndrome (*DHCR7*)<sup>6</sup>, and higher rates of gene-disrupting genetic variants in ASD cases compared with their siblings or controls<sup>7–9</sup>. The set of genes that can be definitively implicated as ASD risk genes has been growing rapidly, and predictive models from studies of single nucleotide variants in sporadic cases estimate that there are likely to be hundreds of genes involved in ASD risk<sup>8,10</sup>. Given the scope of this genetic heterogeneity, understanding the biological aetiologies of ASD and designing broadly applicable treatments has proven challenging.

One robust risk factor for ASD is sex: for every female with ASD in the United States, there are 4.5 affected males<sup>2</sup>, and a male bias in prevalence is consistent across countries and across diagnostic criteria<sup>11</sup>. A multiple threshold liability model has been applied to conceptualize this difference in vulnerability, which posits that a higher minimum liability is required for females to manifest the ASD phenotype as compared with males<sup>12–14</sup>; this is also referred to as the female protective model. There is now evidence from population-wide data, family-level data<sup>14</sup> and at the genetic level<sup>7,9,15</sup> to support the hypothesis from this model that autistic females with ASD carry greater genetic liability. However, we note that a female protective model is not incompatible with the existence of male-specific risk factors, and the molecular mechanisms responsible for either protecting females or potentiating males' vulnerability remain unknown.

Several theories have been proposed regarding sex-differential risk and protective mechanisms, though the evidence supporting each is varied<sup>12</sup>. For example, ASD has been conceptualized as an X-linked disorder<sup>16,17</sup>, and although several X-chromosome genes have been implicated, numerous autosomal genes contribute to risk as well<sup>7,8</sup>; X-chromosome genes alone do not explain the overall male bias<sup>9,18,19</sup>. Imprinted, paternally expressed X-chromosome genes have also been proposed to exert protective effects on females<sup>20</sup>, though such genes have yet to be identified. Beyond genetics, the extreme male brain theory of autism posits that elevated prenatal testosterone exposure increases risk<sup>21</sup>. While recent evidence has linked increased fetal testosterone levels to later ASD diagnoses in a population sample<sup>22</sup>, the molecular and cellular mechanisms that act to link an early hormone exposure to a later ASD phenotype are unknown. It is also plausible that multiple mechanisms, acting independently or interacting with one another, collectively account for the sex bias in ASD risk.

We reasoned that since sex-differential gene expression patterns contribute to the development and function of a sexually dimorphic brain, evaluating genome-wide sex-differential gene expression in neural tissue could elucidate points of overlap with ASD risk genes and related pathways. Here we test two basic hypotheses about the relationship between sexually dimorphic expression and ASD risk genes: (1) ASD risk genes are expressed at different levels in males and females. With differing baseline expression levels, the magnitude of the impact of a disruptive mutation in a risk gene is likely to differ by sex. Accordingly, we expect to observe enrichment of sex-differentially expressed (sex-DE) genes among known ASD risk genes. (2) ASD risk genes are expressed at the same level in males and females, but genes in interacting molecular pathways and/or cellular processes are differentially expressed by sex. In this case, the downstream impact of sex-neutrally expressed ASD risk genes is likely modulated by their interactions with sexually dimorphic

processes. If this were the case, we would expect to observe enrichment of sex-DE genes among gene sets representing processes associated with ASD pathophysiology, but not among ASD risk genes themselves.

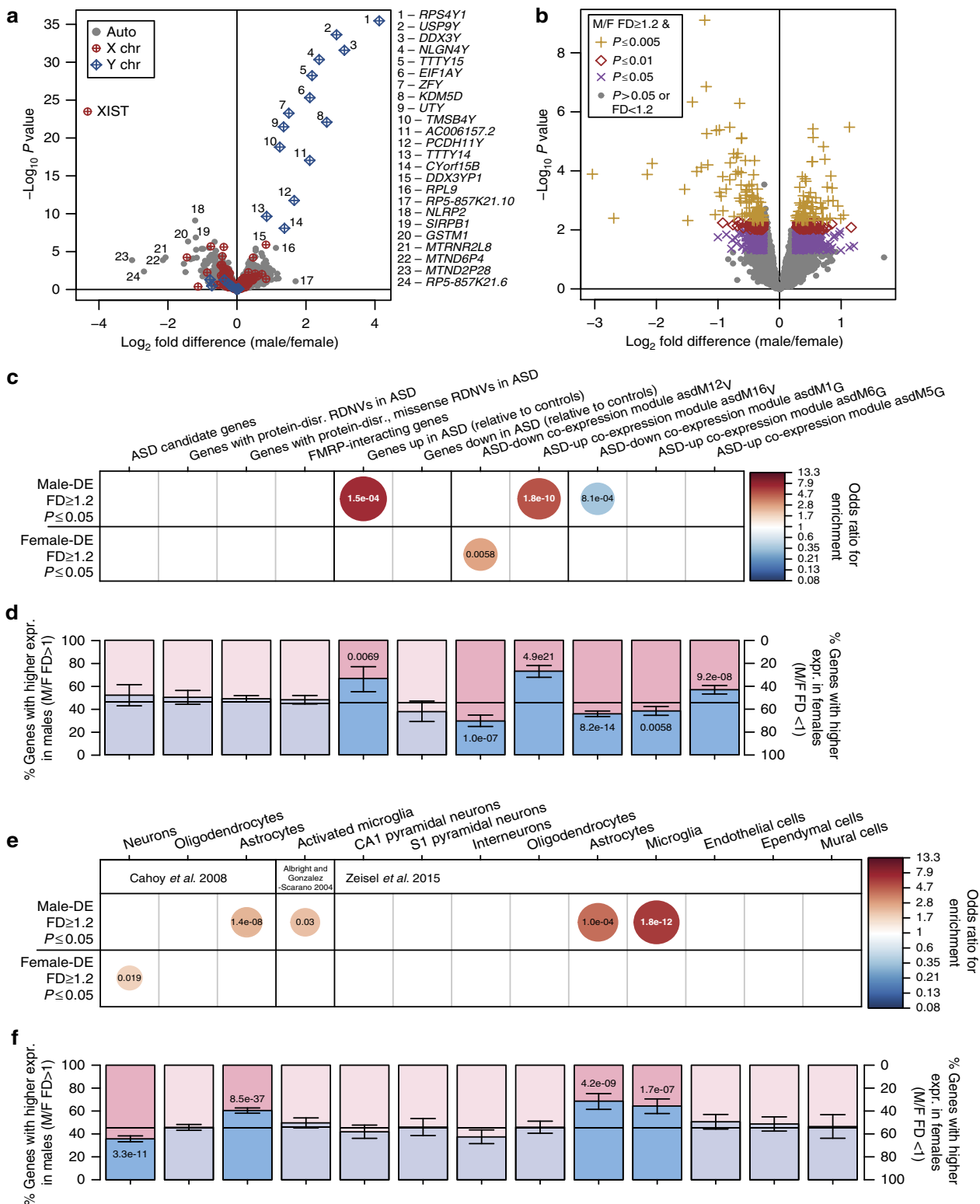
To evaluate these hypotheses, we identify sexually dimorphic gene expression patterns in adult and prenatal human cerebral cortex tissue and test ASD-associated gene sets for enrichment of sex-differential expression. We observe significantly male-biased expression patterns for microglial-function-enriched, ASD-upregulated, co-expression modules from post-mortem ASD brain<sup>23,24</sup> and microglia and astrocyte marker genes<sup>25–27</sup>, but no enrichment of sex-differential expression for ASD risk genes. These patterns are most consistent with our second hypothesis, and suggest that sex differences in biological processes captured by specific ASD-dysregulated co-expression modules, or sexual dimorphisms in cortical microglia and/or astrocytes, may play critical roles in setting males' greater risk for ASD.

## Results

To identify sex-differential gene expression patterns in the human brain, we analysed RNA-sequencing (RNA-seq) gene expression data from human cerebral cortex tissue from the BrainSpan project<sup>28</sup> and from an independent, in-house data set, as well as published array data from prenatal samples<sup>29</sup>. We focused on cerebral cortex, since ASD risk genes are highly expressed in this region in neurotypical brain<sup>30,31</sup>, and examination of post-mortem autistic brain has demonstrated consistent gene expression changes in the cerebral cortex as well<sup>23,24</sup>. To characterize stable sex differences in gene expression, we evaluated samples from adult subjects, and to assess sex-differential expression during early development, when many ASD risk genes are highly expressed, we evaluated samples from prenatal subjects. We then tested multiple ASD risk gene sets, ASD-associated gene sets and cell type markers for enrichment of sex-differential expression<sup>8,23–27,32–34</sup>, to determine whether sexually dimorphic utilization of ASD risk genes, or genes in interacting processes, may contribute to the sex bias in ASD prevalence.

**Sex-differential expression in adult human cortex.** We first analysed RNA-seq data from 58 post-mortem cortex samples from the BrainSpan project<sup>28</sup> for sex-differential gene expression using a linear mixed model. This set of samples was collected from 10 teenage and adult subjects matched for sex, age and brain region (13–40 years; 5 females), and filtered for outliers (Supplementary Table 1).

Sex-differential expression patterns in adult cortex show expected, robust differential expression of Y-chromosome genes and *XIST*, an X-chromosome transcript that initiates X-chromosome inactivation and is only expressed in females (Fig. 1a). We note that all Y-chromosomal transcripts that fail to show male-biased expression are pseudogenes with high sequence similarity to their corresponding genes on other chromosomes. Previous studies of sex-differential expression in human brain have preferentially reported sex-DE genes with pronounced fold differences (FDs), such as those on the X and Y chromosomes<sup>29,35,36</sup>. Beyond these differences driven by sex chromosome copy number, sex-differential expression of autosomal and non-*XIST* X-chromosome transcripts are generally subtle in magnitude<sup>37</sup>, and may reflect sex-specific tuning of molecular pathways. Using a minimum FD magnitude of 1.2, a standard threshold, we identify 186 genes at  $P < 0.005$ , 311 genes at  $P < 0.01$  and 866 genes at  $P < 0.05$  (Fig. 1b and Supplementary Data 1); just 58 genes are sex-DE at a Benjamini–Hochberg adjusted  $P$  value of 0.05, 21 of which are on the sex chromosomes.



**Figure 1 | Microglia and astrocyte markers and genes upregulated in ASD brain tend toward higher expression in adult male brain.** (a) Volcano plot for the differential expression results from all 16,392 transcripts expressed in the adult BrainSpan sample ( $n$  male = 29 samples from 5 subjects,  $n$  female = 29 samples from 5 subjects). (b) Subset of the volcano plot in a for all 15,724 autosomal transcripts. (c) Enrichment for ASD risk genes and ASD-associated gene expression patterns, and (e) neural cell type markers, within male-DE (higher expression in males,  $FD \geq 1.2$ ,  $P \leq 0.05$ ; 439 genes) and female-DE (higher expression in females,  $FD \geq 1.2$ ,  $P \leq 0.05$ ; 427 genes) gene sets by Fisher’s exact test; circle size and colour indicate the odds ratio of all significant overlaps (Bonferroni-adjusted  $P$  value  $< 0.05$ ); overlaid text displays the adjusted  $P$  value for each enrichment. (d) Shifts in the distribution of sex-differential expression direction for genes in each ASD risk or ASD expression sets, and (f) neural cell type markers. Blue and pink bars display the proportions of each gene set that have higher expression in males ( $FD > 1$ ) or females ( $FD < 1$ ); whiskers note 95% confidence intervals; horizontal black lines note the proportion of male- and female-higher genes in the corresponding background gene set. Overlaid text displays significant Bonferroni-adjusted  $P$  values from the binomial test. chr, chromosome; disr., disrupting; expr., expression; F, female; M, male.

The 37 autosomal sex-DE genes, then, are not sexually dimorphic due to copy number differences between males and females. Instead, they may be the regulatory targets of sex steroid hormone receptors or transcription factors from the X or Y chromosomes, or they may be preferentially expressed in a certain cell type that is differentially represented in male and female cortex.

**Over-representation analysis for gene sets of interest.** To determine whether genes associated with ASD risk show sex-differential expression in the human adult cortex, we examined sets of known risk genes from several sources: (a) candidate genes from a manually curated database<sup>32</sup>; (b) genes with rare, *de novo*, protein-disrupting or missense single nucleotide variants in sporadic ASD cases from the Simons Simplex Collection<sup>8</sup>; and (c) FMRP (Fragile X mental retardation protein) binding targets<sup>33</sup> (Supplementary Data 2). Together, these gene sets capture a broad scope of ASD risk genes, including heavily studied candidates, genes with evidence of *de novo* risk variants, and genes from a strongly ASD-associated regulatory network<sup>33</sup>. For each gene set, we first applied a Fisher's exact test to assess its overlap with sex-DE genes defined by specific thresholds ( $FD \geq 1.2$  and  $P \leq 0.05$ , Fig. 1c, Supplementary Data 3;  $FD \geq 1.2$  and  $P \leq 0.01$  or  $P \leq 0.005$ , Supplementary Data 3). To corroborate these results and assess sexually dimorphic shifts in gene expression across the transcriptome without applying arbitrary thresholds, we also used a binomial test to assess each gene set's distribution into two possible outcomes: higher expression in males ( $FD > 1$ , any  $P$  value), or higher expression in females ( $FD < 1$ , any  $P$  value). We found no evidence from either test for significant sex-differential expression of any of these diverse sets of ASD risk genes in the adult cerebral cortex, providing no evidence supporting hypothesis 1 (Fig. 1d and Supplementary Data 3).

Next, we tested gene sets with evidence of ASD-associated expression patterns in adult human cortex<sup>23,24</sup> (Supplementary Data 2), neural cell type marker gene sets<sup>25–27</sup> (Supplementary Data 4) and genes marking cell types of the adaptive immune system<sup>34</sup> (Supplementary Data 5) for sex-differential expression. We include immune cell markers in this analysis, since recent work has suggested complementary roles for T cells and microglia in mediating behaviour, specifically pain responses, in female versus male animals<sup>38</sup>. The ASD-associated gene sets include genes differentially expressed in post-mortem cortex from subjects diagnosed with ASD<sup>24</sup>, and two ASD-associated co-expression modules from this same study (asdM12<sub>V</sub>, downregulated in ASD and enriched for genes with neuronal and synaptic functions; and asdM16<sub>V</sub>, upregulated in ASD and enriched for genes involved in immune and inflammatory responses)<sup>24</sup>. We also tested three ASD-associated co-expression modules from a subsequent RNA-seq study by Gupta *et al.*<sup>23</sup>, which used a larger, but overlapping set of individuals (asdM1<sub>G</sub>, downregulated in ASD, and asdM6<sub>G</sub>, upregulated in ASD, both modules are enriched for neuronal markers and synaptic genes and overlap with the asdM12<sub>V</sub> module<sup>24</sup>; asdM5<sub>G</sub>, which is upregulated in ASD, enriched for M2-state microglial genes, and overlaps with the asdM16<sub>V</sub> module<sup>24</sup>). These DE genes and co-expression modules comprise large gene sets that were coherently altered in samples of ASD cases harbouring different genetic aetiologies. Therefore, these sets are likely to represent the downstream consequences of deleterious variants in risk genes, an upstream or ongoing background of molecular risk for ASD, or a secondary response to alterations in brain function that accompany ASD, any of which might be sexually dimorphic.

In contrast to the ASD risk genes, we observe significant enrichment and depletion for neural cell type markers and gene

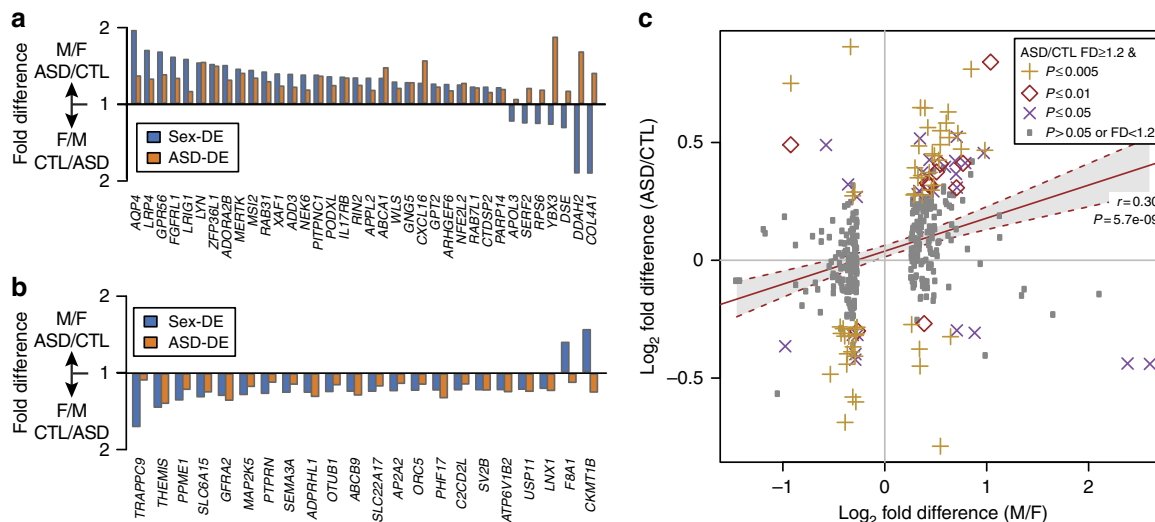
sets dysregulated in post-mortem brain from autistic subjects<sup>24</sup>. Among genes expressed at significantly higher levels ( $FD \geq 1.2$ ,  $P < 0.05$ ) in males than in females (male-DE), we find a sevenfold enrichment of ASD-upregulated genes (7.0-fold, Fisher-adjusted  $P$  ( $P_{\text{Fisher\_adj}}$ ) = 1.5e-04; Fig. 1c and Supplementary Data 3). Among genes upregulated in ASD post-mortem cortex (ASD-up)<sup>24</sup>, 67% show higher expression in males than in females (male-higher, 21% shift from background expectation, binomial test-adjusted  $P$  ( $P_{\text{Binom\_adj}}$ ) = 6.9e-03, Fig. 1d). Genes belonging to the ASD-upregulated module asdM16<sub>V</sub> also show enrichment for significantly male-DE genes (5.2-fold,  $P_{\text{Fisher\_adj}}$  = 1.8e-10, Fig. 1c) and a significant shift towards male-biased expression overall (73% male-higher, 28% shift,  $P_{\text{Binom\_adj}}$  = 4.9e-21, Fig. 1d). The related co-expression module asdM5<sub>G</sub> from Gupta *et al.*<sup>23</sup> also shows this male-biased shift in expression (57% male-higher, 11% shift,  $P_{\text{Binom\_adj}}$  = 9.2e-08, Fig. 1d). It is critical to note that this apparent concordance between sex and ASD status is not the result of a male skew or lack of sex balance in the cases and controls used by either post-mortem ASD study<sup>23,24</sup>, as all samples consisted predominantly of males (Voineagu *et al.*<sup>24</sup>: 4/16 female ASD cases and 1/16 female controls; Gupta *et al.*<sup>23</sup>: 8/32 female ASD cases and 9/41 female controls). Both of these published studies use multiple approaches to show that sex confounds do not drive the observed ASD-associated gene expression changes in brain<sup>23,24</sup>.

The ASD-upregulated gene list asdM16<sub>V</sub>, and asdM5<sub>G</sub> are associated with astrocyte and microglial function; asdM16<sub>V</sub> overlaps significantly with human microglial and astrocyte co-expression modules identified in neurotypical cerebral cortex<sup>24,39</sup> and asdM5<sub>G</sub> is enriched for gene expression signatures from M2 microglia<sup>23</sup>. Consistent with this, we also observe significant enrichments among male-DE genes for two independently generated sets of microglia markers (Albright *et al.*<sup>25</sup>: 2.1-fold,  $P_{\text{Fisher\_adj}}$  = 0.029; Zeisel *et al.*<sup>27</sup>: 6.4-fold,  $P_{\text{Fisher\_adj}}$  = 1.8e-12, Fig. 1e, Supplementary Data 3) and for two independently generated sets of astrocyte markers (Cahoy *et al.*<sup>26</sup>: 2.4-fold,  $P_{\text{Fisher\_adj}}$  = 1.4e-08; Zeisel *et al.*<sup>27</sup>: 4.2-fold,  $P_{\text{Fisher\_adj}}$  = 1.0e-04, Fig. 1e); this is corroborated by significant shifts towards male-biased expression for both astrocyte gene sets and the microglial markers generated by single-cell RNA-seq<sup>27</sup> (Cahoy *et al.*<sup>26</sup>, astrocytes: 60% male-higher, 15% shift,  $P_{\text{Binom\_adj}}$  = 8.5e-37; Zeisel *et al.*<sup>27</sup>, astrocytes: 69% male-higher, 23% shift,  $P_{\text{Binom\_adj}}$  = 4.2e-09; microglia: 64% male-higher, 19% shift,  $P_{\text{Binom\_adj}}$  = 1.7e-07, Fig. 1f).

In contrast, the forebrain neuronal marker gene set<sup>26</sup> is significantly enriched for genes expressed at significantly higher levels in females than in males ( $FD \geq 1.2$ ,  $P < 0.05$ ; female-DE; 1.8-fold,  $P_{\text{Fisher\_adj}}$  = 0.019, Fig. 1e) and shifted towards female-biased expression overall (64% with higher expression in females (female-higher), 9% shift,  $P_{\text{Binom\_adj}}$  = 3.3e-11, Fig. 1f). Similarly, genes in the ASD-downregulated, neuronal and synaptic module asdM12<sub>V</sub> show enrichment for female-DE genes (2.7-fold,  $P_{\text{Fisher\_adj}}$  = 5.8e-03, Fig. 1c), while the ASD-downregulated module asdM1<sub>G</sub> shows depletion of male-DE genes (0.37-fold,  $P_{\text{Fisher\_adj}}$  = 8.1e-04, Fig. 1c); both show a significant shift towards female-biased expression overall (asdM12<sub>V</sub>: 70% female-higher, 16% shift,  $P_{\text{Binom\_adj}}$  = 1.0e-07; asdM1<sub>G</sub>: 64% female-higher, 10% shift,  $P_{\text{Binom\_adj}}$  = 8.2e-14, Fig. 1d). No T- or B-cell modules showed any significant enrichment with sex-DE genes or any significant shift towards male- or female-biased expression (Supplementary Fig. 1a).

#### Comparison of sex- and ASD-differential expression patterns.

We were particularly interested in the sex-DE enrichment in the ASD-associated co-expression modules from post-mortem brain, asdM16<sub>V</sub> and asdM12<sub>V</sub> (ref. 24). Previous studies have



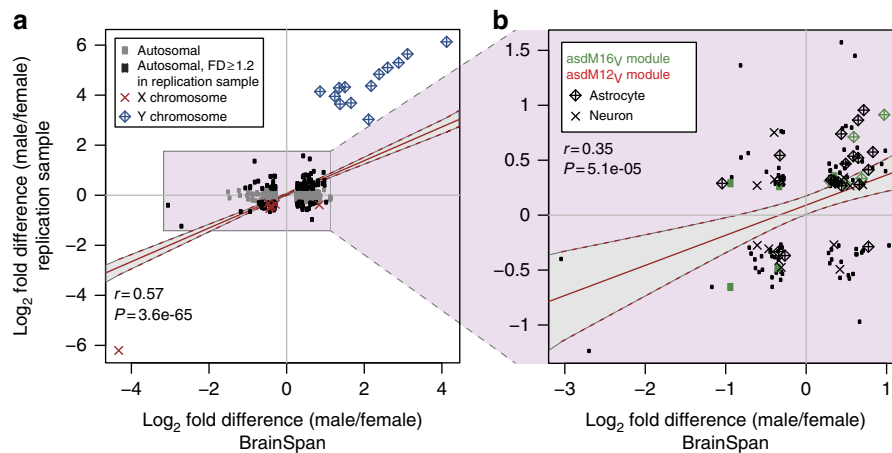
**Figure 2 | Sex-differential expression parallels gene expression patterns in ASD brain.** Bar plots show FDs by sex and ASD status of sex-DE genes ( $FD \geq 1.2$  and  $P < 0.05$ ) that also belong to ASD-associated co-expression modules<sup>24</sup>. **(a)** Module  $asdM16_v$ , which is significantly upregulated in ASD and enriched for genes involved in the inflammatory response and immune system functions, and **(b)** module  $asdM12_v$ , which is significantly downregulated in ASD and enriched for genes with neuronal and synaptic functions. **(c)** FDs of sex-DE genes ( $FD \geq 1.2$ ,  $P < 0.05$ , 374 genes) by sex and status. Best-fit line and its 95% confidence interval are shown. CTL, control from ASD expression study; F, female; M, male.

demonstrated that co-expression modules frequently correspond to neural cell types and coherent biological functions<sup>39,40</sup>; thus, sex-DE enrichments in these modules may most directly implicate pathways involved in sex-biased ASD risk. Strikingly, we observe that all 36 sex-DE genes that are also members of  $asdM16_v$  are expressed at higher levels in ASD than in control cortex<sup>24</sup>, 29 of which are also expressed at higher levels in neurotypical male than in female cortex (Fig. 2a). We also observe that all of the 22 sex-DE genes in  $asdM12_v$  are expressed at lower levels in ASD than in control cortex, and 20 of these are also expressed at higher levels in females compared with males (Fig. 2b). In short, we find concordant directionality of differential expression by sex and ASD status, with  $asdM16_v$  sex-DE genes expressed at higher levels in ASD and typical males and  $asdM12_v$  sex-DE genes expressed at higher levels in controls and typical females.

To assess the extent of these parallel expression patterns across the transcriptome, we compared the FDs from sex-DE genes with FDs observed in the differential expression analysis of ASD cortex<sup>24</sup>. For the 374 sex-DE genes ( $FD \geq 1.2$ ,  $P \leq 0.05$ ) that were also tested in the ASD-DE analysis, we observe a highly significant positive correlation between the sex-differential and ASD-differential FDs ( $r = 0.30$ ,  $P = 5.7e-09$ , Fig. 2c), as well as for the subset of genes differentially expressed at  $FD \geq 1.2$  and  $P \leq 0.05$  in both comparisons ( $r = 0.29$ ,  $P = 5.7e-04$ ,  $N = 90$  genes). We note that among the 47 genes expressed at significantly higher levels in both males and in ASD, 22 are  $asdM16_v$  members<sup>24</sup> and 19 are astrocyte markers<sup>26</sup> (16 belong to both sets). Among the 25 genes expressed at significantly higher levels in both females and in unaffected controls, 11 are  $asdM12_v$  members<sup>41</sup> and 12 are neuron markers<sup>14</sup> (5 belong to both sets). These results suggest that the parallels in sex-differential and ASD-differential expression direction extend beyond  $asdM16_v$  and  $asdM12_v$ , and that this pattern may be more broadly indicative of sexually dimorphic transcriptomes originating from specific neural cell types.

**Independent replication of adult sex-differential expression.** To validate these observations, we next analysed RNA-seq data from an independent sample of adult cortex tissue (16–56 years,

Supplementary Table 1) for sex-differential expression patterns. Using this smaller data set, we identify 65 genes at  $P < 0.005$ , 98 genes at  $P < 0.01$  and 519 genes at  $P < 0.05$  (Supplementary Fig. 2a,b and Supplementary Data 1). Overall, sex-DE genes from the BrainSpan data were expressed at comparable levels and with positively correlated fold differences in this independent sample (expression levels:  $r = 0.78$ ,  $P = 2.9e-150$ ;  $\log_2(FD)s$ ,  $r = 0.57$ ,  $P = 3.6e-65$ ; Fig. 3a). We observed no over-representations of ASD risk genes among significantly sex-DE genes ( $FD \geq 1.2$ ,  $P < 0.05$ ) from these data (Supplementary Fig. 2d and Supplementary Data 6); however, FMRP targets were more likely to show higher expression in females than in males (62% female-higher, 13% shift,  $P_{Binom\_adj} = 2.9e-10$ , Supplementary Fig. 2d). Oligodendrocyte markers and endothelial cell markers also showed shifts towards female-biased expression (Cahoy *et al.*<sup>26</sup> oligodendrocytes: 56% female-higher, 9% shift,  $P_{Binom\_adj} = 1.5e-11$ ; Zeisel *et al.*<sup>27</sup> oligodendrocytes: 59% female-higher, 15% shift,  $P_{Binom\_adj} = 5.8e-07$ ; endothelial cells<sup>27</sup>: 55% female-higher, 11% shift,  $P_{Binom\_adj} = 5.7e-03$ , Supplementary Fig. 2f). We also found depletion of male-DE genes in  $asdM1_G$  (0.26-fold,  $P_{Fisher\_adj} = 0.027$ , Supplementary Fig. 2c), but enrichments for  $asdM16_v$  genes (4.7-fold,  $P_{Fisher\_adj} = 2.8e-03$ ; 64% male-higher, 11% shift,  $P_{Binom\_adj} = 5.2e-03$ , Supplementary Fig. 2c,d),  $asdM6_G$  genes (58% male-higher, 7% shift,  $P_{Binom\_adj} = 0.014$ , Supplementary Fig. 2d), astrocyte marker genes (Cahoy *et al.*<sup>26</sup>: 61% male-higher, 10% shift,  $P_{Binom\_adj} = 1.7e-16$ ; Zeisel *et al.*<sup>27</sup>: 69% male-higher, 18% shift,  $P_{Binom\_adj} = 6.6e-06$ , Supplementary Fig. 2f) and ependymal cell markers (60% male-higher, 9% shift,  $P_{Binom\_adj} = 0.027$ , Supplementary Fig. 2f). Again, we find no sex-differential expression signal in T- and B-cell modules (Supplementary Fig. 1b and Supplementary Data 6). Of the 164 genes expressed more highly in males with a  $FD \geq 1.2$  in both adult data sets, 24 are  $asdM16_v$  members (8.8-fold enrichment,  $P = 5.5e-14$ ) and 70 are astrocyte markers (10.3-fold enrichment,  $P = 3.3e-35$ ; 14 of these genes are belong to both sets), further supporting the characterization of astrocytic genes and ASD-upregulated, microglial-function-enriched  $asdM16_v$  genes, as displaying male-biased expression in the adult human cortex (Fig. 3b).



**Figure 3 | Male-biased expression of astrocyte and asdM16<sub>v</sub> module genes in an independent sample.** (a) FD for all sex-DE genes from the BrainSpan sample also tested in the replication set (BrainSpan  $FD \geq 1.2$ ,  $P < 0.05$ , 733 genes; grey points mark genes with  $FD < 1.2$  in the replication sample), and for (b) autosomal sex-DE genes from BrainSpan with  $FD \geq 1.2$  in the replication sample (129 genes). Best-fit lines and 95% confidence intervals are shown on plots.

**Sex-differential expression in prenatal human cortex.** One possible explanation for the lack of enrichment of ASD risk genes among sex-DE genes from adult brain is that ASD risk genes may be acting primarily during fetal brain development<sup>30,31</sup>. To address this, we next analysed array expression data from 86 neocortex samples from 8 subjects (4 females) between 16–22 post-conception weeks<sup>29</sup> (Supplementary Table 1). This prenatal epoch follows the mid-gestation peak in testosterone secretion from males' differentiated testes<sup>42,43</sup> and thus captures a stage during which steroid hormones are likely to be exerting organizational effects.

We find 306 sex-DE genes at  $P < 0.005$ , 440 sex-DE genes at  $P < 0.01$  and 1,037 sex-DE genes at  $P < 0.05$ , with 96 genes sex-DE at an adjusted  $P < 0.05$  (Fig. 4a,b and Supplementary Data 1). Results from over-representation analysis for the gene sets of interest within these prenatally sex-DE genes mirror those in the adult data: there are no significant enrichments for ASD risk gene sets, and FMRP interactors are significantly depleted within male-DE genes (0.33-fold,  $P_{\text{Fisher\_adj}} = 8.0e-04$ , Fig. 4c, Supplementary Data 7).

As in adult brain, genes expressed more highly in prenatal male cortex show robust enrichments for post-mortem cortex ASD-upregulated genes (4.6-fold,  $P_{\text{Fisher\_adj}} = 3.3e-03$ , Fig. 4c), asdM16<sub>v</sub> genes (4.0-fold,  $P_{\text{Fisher\_adj}} = 2.6e-09$ ; 67% male-higher, 16% shift,  $P_{\text{Binom\_adj}} = 6.2e-05$ , Fig. 4c), asdM5<sub>G</sub> genes (2.9-fold,  $P_{\text{Fisher\_adj}} = 5.1e-09$ ; 61% male-higher, 13% shift,  $P_{\text{Binom\_adj}} = 4.2e-07$ , Fig. 4c,d), microglia markers (Albright *et al.*<sup>25</sup>: 2.1-fold,  $P_{\text{Fisher\_adj}} = 2.8e-03$ ; 60% male-higher, 11.4% shift,  $P_{\text{Binom\_adj}} = 1.2e-04$ ; Zeisel *et al.*<sup>27</sup>: 2.9-fold,  $P_{\text{Fisher\_adj}} = 0.027$ , Fig. 4e,f), endothelial markers (3.7-fold,  $P_{\text{Fisher\_adj}} = 3.2e-06$ , Fig. 4e) and a significant depletion of forebrain neuron marker genes (0.41-fold,  $P_{\text{Fisher\_adj}} = 2.3e-05$ ; 58% female-higher, 7% shift,  $P_{\text{Binom\_adj}} = 1.3e-04$ , Fig. 4e,f). As in adult, T- and B-cell modules show no enrichments or depletions of sex-DE genes, nor any sex-skewed expression shifts (Supplementary Fig. 1c and Supplementary Data 7).

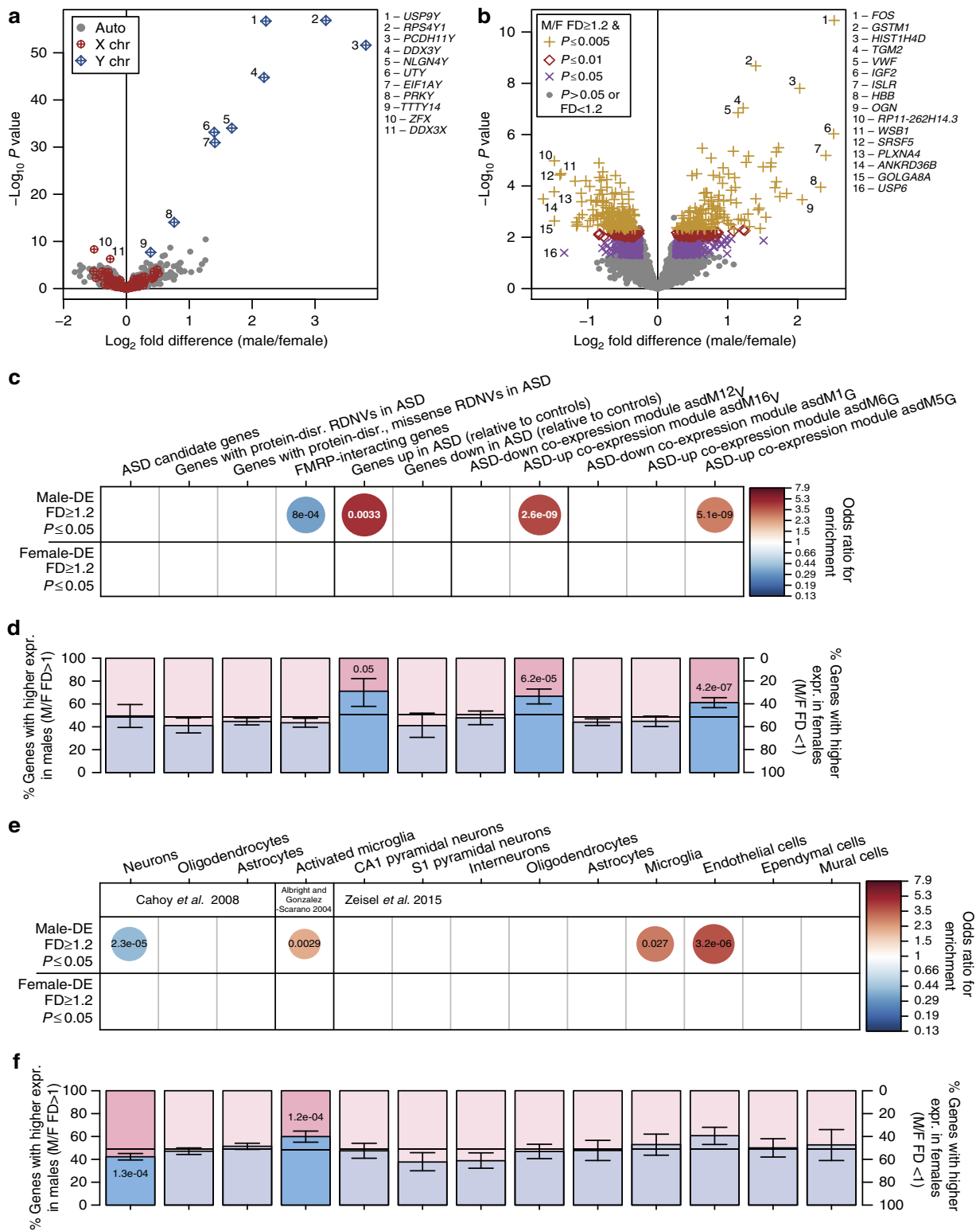
Taken together, these findings fail to support the hypothesis that ASD risk genes are differentially expressed by sex in the human cerebral cortex. Rather, we observe that sex-DE genes are enriched among specific classes of changes observed in post-mortem cortex from ASD patients, which is indicative of shared biology between ASD and typical sexual dimorphisms. These patterns of sex-DE enrichment follow what is observed in ASD cortex, with higher expression of microglial and astrocyte genes

(exemplified by asdM16<sub>v</sub> and asdM5<sub>G</sub>) in the more susceptible male population, and higher expression of a cohort of synaptic genes (exemplified by asdM12<sub>v</sub>), in the more protected, female population.

## Discussion

Autism risk is sexually dimorphic, and genetic studies are most consistent with a female protective effect. To begin to explore potential mechanisms mediating sex-differential genetic risk, we tested the hypothesis that ASD risk genes would show sexually dimorphic expression in the cerebral cortex, a region that plays a major role in complex human cognitive phenotypes, including social behaviour, and where many ASD genes are expressed<sup>30,31</sup>. We observe both in adult and fetal brain that while some ASD risk genes do show sex-differential expression, overall this is not beyond what would be expected by chance. Furthermore, these sex-neutral expression patterns are evident in the mid-fetal cortex, a spatiotemporal point of convergence for ASD risk genes' expression<sup>30,31</sup>.

Instead, we find support for a second hypothesis: that genes whose expression patterns are altered in post-mortem autistic brain, which highlight downstream or interacting pathways, but not ASD risk genes themselves, are differentially expressed by sex. The convergence of male-DE genes with these downstream pathways, but not with risk genes or neuronal markers, is consistent with the evidence presented by Voineagu *et al.*<sup>24</sup>: asdM16<sub>v</sub> is upregulated in ASD brain and enriched for immune system function and inflammatory response genes, but not for common ASD susceptibility variants. The asdM16<sub>v</sub> module is therefore interpreted as representing a secondary, likely downstream effect of genetically causal perturbations. Whether a secondary effect of genetic variants or a result of independent mechanisms, it may be that expression of this gene set beyond a certain threshold is detrimental to neural development and function. Our findings show that typical males land closer to this putative threshold than females do, potentially implicating these genes expressed at higher levels in the male and ASD cortex in mechanisms driving male-biased ASD risk. Furthermore, to the extent that asdM16<sub>v</sub> upregulation contributes to ASD pathophysiology, this module may serve as a useful target for pharmacological treatments that could modulate the effects of heterogeneous risk variants acting upstream.



**Figure 4 | Microglia and astrocyte markers and genes upregulated in ASD brain tend toward higher expression in prenatal male brain.** (a) Volcano plot for all 9,889 transcripts expressed in the prenatal sample ( $n$  male = 43 samples from 4 subjects,  $n$  female = 43 samples from 4 subjects). (b) Subset of the plot in a for the 9,532 autosomal transcripts. (c) Enrichment for ASD risk genes and ASD-associated gene expression patterns, and (e) neural cell type markers, within male-DE (509 genes) and female-DE (528 genes) gene sets by Fisher’s exact test; circle size and colour indicate the odds ratio of all significant overlaps (Bonferroni-adjusted  $P$  value  $<$ 0.05); overlaid text displays the adjusted  $P$  value for each enrichment. (d) Shifts in the distribution of sex-differential expression direction for genes in each ASD risk or ASD expression sets, and (f) neural cell type markers. Blue and pink bars display the proportions of each gene set that have higher expression in males (FD $>$ 1) or females (FD $<$ 1); whiskers note 95% confidence intervals; horizontal black lines note the proportion of male- and female-higher genes in the corresponding background gene set. Overlaid text displays significant Bonferroni-adjusted  $P$  values from the binomial test. chr, chromosome; disr., disrupting; expr., expression; F, female; M, male.

Alternatively or concordantly, it may also be the case that ASD-protective genes are expressed at a higher level in females. Genes in the neuronal- and synapse function-enriched asdM12<sub>v</sub> module may represent such a protective gene set, as these genes are downregulated in ASD post-mortem cortex<sup>24</sup>. Genes belonging to asdM12<sub>v</sub> do show higher expression in adult female cortex from the BrainSpan sample (Fig. 1c,d), though neither the replication adult sample nor the prenatal sample show this same female skew. Alternatively, the relatively sex-neutral expression of asdM12<sub>v</sub> genes in contrast to asdM16<sub>v</sub> may simply reflect the absence of a particular ASD pathophysiological process in neurotypical samples.

We also observed a correlation in the relative FDs of sex-DE genes and ASD-DE genes that extends beyond the asdM16<sub>v</sub> module, such that genes expressed more highly in typical male cortex also tend to be more highly expressed in ASD cortex, while genes more highly expressed in typical female cortex also tend to be expressed at lower levels in ASD cortex<sup>24</sup>. Again, we emphasize this agreement in fold change direction between sex and ASD status is not likely the result of a male-confounded analysis by Voineagu *et al.*<sup>24</sup> or Gupta *et al.*<sup>23</sup>, as cases and controls were sex balanced, and any potential sex effects were also regressed out in those studies. Interestingly, the gene expression patterns we observe fit squarely with hypotheses derived from both the extreme male brain theory and from the concept of female protective factors in ASD: male-DE (ASD-upregulated) genes could act to increase risk, while female-DE (ASD-downregulated) genes could have protective functions. However, we caution that FDs do not indicate the underlying mechanism of gene regulation (that is, active upregulation in males can return the same result as repression in females). We cannot yet conclude that parallel expression differences between the sexes and in ASD are derived from the same regulatory mechanisms, but this work provides a starting point for their study. Additional experiments that include sufficient numbers of cases and controls of both sexes will be necessary to tease apart how sex and ASD status interact to affect gene expression, for example, to determine whether male-DE genes are expressed at even higher levels in ASD cases.

Astrocyte and microglial markers are also reliably enriched among male-DE genes in both adult and prenatal cortex. We recognize two possible explanations for this: (1) a greater proportion of astrocytes and microglia to neurons in male cortex results in higher measured gene expression in males; or (2) males' cortical astrocytes and microglia are more transcriptionally active than females'. Aside from early, tenuous observations of greater cortical neuron density in males<sup>44</sup>, sex differences in the cellular composition of human cortex have not been sufficiently characterized to determine if there is a greater number of certain glial subtypes in male cortex. In contrast, there is clear evidence for sexually dimorphic astrocyte morphology in animals, particularly in hypothalamic nuclei, with male astrocytes showing a greater number of longer and more branched processes<sup>45–47</sup>. In addition, sex steroid hormones may play a role in astrocyte differentiation and regulation, as oestrogen receptors are expressed in hypothalamic and hippocampal astrocytes<sup>48,49</sup>, and female animals exposed to testosterone neonatally show male-typical astrocyte morphology<sup>46</sup>. Given the participation of astrocytes in the modulation of neurotransmission<sup>50,51</sup> and of microglia and astrocytes in synapse formation and function<sup>52–55</sup>, it is plausible that sexual dimorphisms in astrocyte and/or microglia number or function would lead to sex-differential effects on neuronal connectivity<sup>56</sup>. It is also interesting to consider that numerous ASD risk genes function at the synapse<sup>7,9,57–59</sup>, and that as a third synaptic component along with pre- and postsynaptic neurons, sexually dimorphic astrocytes and/or

microglia are well positioned to broadly influence the effects of upstream, heterogeneous risk variants.

We also acknowledge several limitations of our study. First, our analyses utilize gene expression data from the human cerebral cortex, a region that other studies have indicated is a convergent area where many ASD risk genes likely exert their action<sup>30,31</sup>. However, other brain regions not sampled in our data sets, such as the hypothalamic nuclei, are known to be robustly sexually dimorphic<sup>60</sup>. It may be that ASD risk genes are differentially expressed by sex in these sexually dimorphic regions, or that it is the connections between these regions and other parts of the brain where the sex-differential risk-modulatory mechanisms act (as opposed to purely within-cortex interaction at the level of molecular pathways and local cell–cell connections). Assessment of additional brain regions, including hypothalamic nuclei and other subcortical structures, in sufficiently powered test and replication sets will be useful for addressing this important question, once such data are available.

The mechanism(s) of the sex-biased expression observed here are not known, but present significant opportunities for understanding ASD's sex-biased pathogenesis. For example, transcription of sex-DE genes identified here may be regulated by sex-differential factors, such as transcription factors from the sex chromosomes, or by sex steroid receptor binding to hormone response elements upstream from transcription start sites. Chromatin immunoprecipitation experiments in human neural tissue coupled with functional assays in model systems will be needed to thoroughly characterize binding sites for these regulatory elements throughout the genome.

Alternatively, the sex differences we have observed may not result from differential transcription regulation, but instead from sex differences in the proportions of different neural cell types within male and female tissue. The overlap between male-DE genes and astrocyte and microglial marker genes is consistent with this possibility, although the processes that lead to sex differences in cell type proliferation or differentiation are unclear. Quantitative histological studies of male and female human brain will be critical to characterize the extent to which cell type composition differs between the sexes, and neurodevelopmental follow-up will be required to demonstrate how these differences arise.

The extent to which the convergent transcriptional changes previously observed in ASD cortex<sup>23,24</sup> represent causal aetiologies versus consequences of brain dysfunction also remains to be delineated<sup>61</sup>. In the case of ASD-associated co-expression modules enriched for risk variants, such as asdM12<sub>v</sub>, enrichment of genes carrying ASD risk variants provides evidence for a causal, primary role in ASD aetiology. However, ASD-associated co-expression modules that lack enrichment for genetic risk variants, such as asdM16<sub>v</sub>, may be linked to ASD in one of several ways: (1) They reflect direct downstream consequences of variants in, or dysregulation of, primary, 'causal' genes (for example, asdM12<sub>v</sub>) that actively participate in the pathophysiological processes that shape and maintain an autistic-functioning brain; (2) they represent risk pathways that are independent of primary risk genes, for example, consequences of an environmental exposure, that also actively participate in ASD's aetiology; or (3) they exemplify consequences of living with ASD to the brain, which may modify the presentation of ASD or its comorbidities, but that do not have a role in ASD's aetiology. For example, the gene expression signature captured by non-variant-enriched, co-expression module asdM16<sub>v</sub>, which is upregulated in post-mortem adult autistic cortex, may reflect the impact of chronic social stress experienced across the lifespan. If male and female brains respond differently to stress<sup>62</sup>, then the concordance we observe between these gene expression signatures may



be tangential to ASD's primary pathophysiology. Though, one would expect that such stress would be cumulative over the lifespan, so that a stress-driven gene expression profile would increase with experience and age. Neither of the ASD-associated co-expression modules is significantly correlated with age<sup>23,24</sup>, making this explanation less likely. Nevertheless, extensive characterization of animal models of ASD genetic risk variants, including gene expression patterns, may help to determine whether expression signatures like these are purely consequential, or if they are evident during early development and play a role in ASD aetiology.

In any case, the sex-DE patterns that we observe in multiple, independent samples demonstrate that currently identified ASD risk genes are not innately sex-differentially expressed in the developing or adult human cerebral cortex. Instead, it is the molecular, cellular or circuit-level context in which ASD risk genes operate that is likely responsible for modulating ASD liability in a sex-differential manner. This notion fits with the remarkable consistency of the male bias in ASD prevalence, despite ongoing evidence of considerable genetic heterogeneity in ASD. Study of astrocyte–neuronal and microglia–neuronal synapse interactions, and study of the causes and functional consequences of higher expression of astrocyte and microglial genes in male brain and synaptic genes in female brain, are needed to determine how sex-differential functioning of these pathways impact sexually dimorphic risk for a variety of neurodevelopmental disorders, including ASD.

## Methods

**Adult BrainSpan data set.** Developmental RNA-seq data from the BrainSpan project<sup>28</sup> summarized to Gencode 10 (ref. 41) gene-level reads per kilobase million mapped reads (RPKM) were used. Data were normalized for GC content, using conditional quantile normalization<sup>63</sup>, batch-corrected for processing site using ComBat<sup>64</sup> and log<sub>2</sub>-transformed (log<sub>2</sub>(RPKM + 1)). Samples from regions of the frontal, temporal and parietal cortex from subjects aged 13–40 years, with RNA integrity number (RIN) of at least 8.0, were retained for this analysis. Non-expressed genes defined as having a log<sub>2</sub>-transformed RPKM expression level of < 1 in more than 50% of all male or female samples were removed. Outlier samples were identified by evaluating inter-sample correlations and hierarchical clustering, first within each sex, and then on the full data set (samples of both sexes considered together); samples with inter-sample correlations > 2.5 s.d.'s from the mean were removed. The data were then quantile normalized to mitigate the effects of systematic differences in the distribution of expression levels across samples on the differential expression analysis results. After gene filtering and outlier removal, 72 samples (29 from males and 43 from females) and 16,843 expressed genes remained. To balance the male and female sample sets, we then matched a subset of the female samples to the male samples by age and brain region, and again filtered for expressed genes. The final data set consisted of 58 samples from 10 subjects (29 samples from 5 subjects of each sex; Supplementary Table 1) and 16,392 genes (Supplementary Data 1).

**Adult replication data set.** The adult replication sample comprised seven frontal and temporal cortex samples from five male individuals matched to seven cortical samples from five female individuals with no history of neuropsychiatric or neurological diagnoses. Male and female samples were matched for age, post-mortem interval (PMI) and brain region. These samples were a subset of a larger set acquired from the Harvard Brain and Tissue Bank and the Eunice Kennedy Shriver National Institute for Child Health and Human Development Brain and Tissue Bank for Developmental Disorders following the tissue acquisition policies of the respective brain banks.

Approximately 100 mg of tissue from each sample was dissected on dry ice; RNA isolations were performed using the miRNeasy kit (Qiagen), RNA quality was quantified using the RIN<sup>65</sup> and ribosomal RNA was depleted using the Ribo-Zero Gold kit (Epicentre). Remaining RNA was then size selected with AMPure XP beads (Beckman Coulter) and re-suspended, and sequencing libraries with indexed adaptors were prepared according to the Illumina TruSeq protocol. Fifty base-pair, paired-end reads were generated on a HiSeq 2000/2500, with 24 samples pooled per lane.

Reads were mapped to the human reference genome (hg19) using Gencode v18 annotations with TopHat2, allowing for up to 10 multiple mappings per read<sup>41,66</sup>. Output BAM files were filtered to ensure that every read had a valid pair, resulting in only paired-end reads (fragments) being used for downstream analyses. Transcript levels were quantified using Gencode v18 gene models at the union gene

model level using HT-seq Counts<sup>67</sup>. We normalized these data for GC content biases using the cqn package in R (ref. 63), which resulted in log<sub>2</sub>(normalized FPKM) values, and ensured that there were no sample outliers with a summed sample correlation Z-score > 2 (ref. 68). Genes not expressed in the subset of neurotypical samples selected for these analyses were identified (log<sub>2</sub>-transformed FPKM expression level < 1 in more than 50% of all male or female samples) and removed. Outlier samples with inter-sample correlations > 2.5 s.d.'s from the mean, either within each sex or within the complete sample set, were removed. The expression data for the remaining samples were then quantile normalized. The final filtered data set consisted of 13 samples from 10 subjects (Supplementary Table 1), and 19,354 genes (Supplementary Data 1).

**Prenatal data set.** Exon array data analysed by Kang *et al.*<sup>29</sup> and downloaded from the Gene Expression Omnibus (GSE25219) were used for the assessment of prenatal gene expression. Only samples from subjects between 16 and 22 post-conception weeks from the frontal, temporal and parietal cortex, and with RIN of 8.0 or greater were used in this stage of analysis.

Probe set IDs from the downloaded data were matched to Ensembl Gene IDs from Gencode 10 using the biomaRt function in R. Non-expressed genes were defined as those genes with a log<sub>2</sub>-transformed median probe set intensity of < 6 in > 80% of male or female samples within the selected sample set, and were removed. Outlier samples were detected by evaluating inter-sample correlations and hierarchical clustering within each sex and within the complete data set, and samples with inter-sample correlation > 2.5 s.d.'s from the mean were removed. Expression data for the remaining samples were quantile normalized. After these processing steps, 142 samples (43 from males and 99 from females) and 9,985 expressed genes remained. To balance the number of male and female samples for analysis, we then matched a subset of the female samples to the male samples on age and brain region, and again filtered for expressed genes. The final data set consisted of 86 samples from 8 subjects (43 samples from 4 subjects of each sex; Supplementary Table 1), and 9,889 genes (Supplementary Data 1).

**Differential expression analysis.** Differential expression analyses for all data sets were performed using a linear mixed model and Bayesian *t*-tests as implemented in LIMMA<sup>69</sup>, a robust method for analysing small samples. For all analyses, the main contrast in the regression model was sex, and subject was included as a random effect to account for the non-independence of samples from the same individual brain (that is, different brain regions). Covariates were included, where available, in the model as fixed effects: RIN, age, PMI, cortical lobe and pH for the adult BrainSpan data; RIN, age, PMI and cortical lobe for the replication data; and RIN, age, PMI, cortical lobe and pH for the prenatal data. Per-subject average values were substituted for missing PMI and pH information. Genes with a FD magnitude of 1.2 or greater and an unadjusted *P* value of 0.005, 0.01 or 0.05 were called as differentially expressed by sex (sex-DE; Supplementary Data 1).

**Annotation gene sets.** ASD candidate genes were selected from the SFARI Gene Autism Database<sup>32</sup> and filtered to those genes classified as syndromic (category S) or with evidence levels between 1–4 for a set of 153 genes. To also investigate a risk gene set based on evidence from an unbiased, experimental screen, we used the set of genes with rare *de novo* variants (RDNVs) in sporadic ASD cases from exome sequencing of families from the Simons Simplex Collection<sup>8</sup>. All genes with RDNVs in autistic probands were compiled and classified as protein-disrupting (nonsense, splice site or frameshift mutations), missense or silent variants. Here we use the gene set with protein-disrupting variants (353 genes) and the expanded gene set with either protein-disrupting or missense variants (1,771 genes). We also evaluated the set of transcripts that are binding targets of FMRP, a functionally defined gene set that is enriched for ASD risk genes<sup>70</sup>. Here we tested the 725 human orthologues of FMRP target genes as characterized in mouse brain<sup>33</sup> (Supplementary Data 2).

Seven gene sets with ASD-associated expression patterns in adult human cortex were also tested. From a 2011 study by Voineagu *et al.*<sup>24</sup>, we tested genes expressed at significantly (1) higher (FD ≥ 1.2, *q*-value < 0.1, 88 genes) or (2) lower levels in ASD than in control cortex (125 genes); (3) ASD-associated, unsigned co-expression modules from weighted gene co-expression network analysis, asdM12<sub>V</sub> (downregulated in ASD, 414 genes); and (4) asdM16<sub>V</sub> (upregulated in ASD, 361 genes). From a 2014 study by Gupta and colleagues<sup>23</sup> that used a larger, but overlapping, sample of individuals as Voineagu *et al.*<sup>24</sup>, but which analysed BA19, a different cortical region, we tested several ASD-associated weighted gene co-expression network analysis modules, including (5) asdM1<sub>G</sub> (downregulated in ASD, 1,643 genes), (6) asdM6<sub>G</sub> (upregulated in ASD, 667 genes) and (7) asdM5<sub>G</sub> (upregulated in ASD, 759 genes) (Supplementary Data 2).

Since gene expression patterns in the brain often reflect the transcriptional activity of distinct neural cell types<sup>39,40</sup>, we also culled gene sets marking different cell types from several studies in the literature. From a microarray study of sorted cell types in mouse forebrain<sup>26</sup>, we tested markers for neurons (1,489 genes), oligodendrocytes (1,626 genes) and astrocytes (1,968 genes). From a microarray study of human microglial cells in culture<sup>25</sup>, we tested genes with at least a twofold upregulation in activated, microglia-enriched, mixed glial cultures (~60% microglia) (773 genes). From a single-cell RNA-seq study of mouse somatosensory

cortex and hippocampus<sup>27</sup>, we tested markers for CA1 pyramidal neurons (355 genes), S1 pyramidal neurons (235 genes), interneurons (319 genes), oligodendrocytes (397 genes), astrocytes (216 genes), microglia (368 genes), endothelial cells (312 genes), ependymal cells (402 genes) and mural cells (pericytes and vascular smooth muscle cells, 137 genes) (Supplementary Data 4).

Recent evidence has also suggested complementary roles for microglia and immune cells in mediating behaviour in male versus female animals<sup>38</sup>. To evaluate whether adaptive immune cells contribute to the sex-differential gene expression that we observe in the brain, we tested co-expression modules generated by the Immunological Genome Project from mouse haematopoietic cells that are associated with T- and B-cell lineage regulator genes<sup>34</sup>; T-cell module 18 (106 genes); gamma delta T-cell module 65 (30 genes); gamma delta, natural killer T and T4 module 56 (58 genes); alpha beta T-cell module 57 (52 genes); B-cell module 33 (125 genes); B-cell module 61 (38 genes); and B-cell module 76 (15 genes); Supplementary Data 5).

For the enrichment analyses, we used the following gene sets as background: for SFARI candidate genes, genes with RDNVs in ASD cases and activated microglial markers, we used expressed genes with a Gencode biotype annotation of 'protein coding'; for gene sets with ASD-associated expression patterns and for immune cell markers, we used expressed genes also tested in the corresponding source publication<sup>23,24,34</sup>, and for FMRP target genes<sup>33</sup> and cell type markers from Cahoy *et al.*<sup>26</sup> and Zeisel *et al.*<sup>27</sup>, we used the genome-wide set of one-to-one human-mouse orthologues as background (Supplementary Data 8).

For all background and gene sets of interest, gene identifiers were converted to Ensembl Gene IDs in Gencode using the biomaRt package in R to allow for unambiguous comparisons between genes from different data sources. Gene set membership, background set membership and sex-DE categorization for all expressed genes tested for differential expression in the three data sets are recorded in Supplementary Data 9.

**Over-representation analysis.** To evaluate the enrichment of gene sets of interest among sex-DE genes, we applied a two-sided Fisher's exact test separately to sex-DE gene sets expressed more highly in males (male-DE) and in females (female-DE) at three *P* value thresholds (unadjusted  $P < 0.005$ , 0.01 and 0.05, Supplementary Data 3, 6 and 7) in R. *P* values from the Fisher's exact tests run within each data set were adjusted for 62 tests performed against the 31 gene sets of interest (4 ASD risk gene sets, 7 ASD expression gene sets, 13 neural cell type marker sets and 7 adaptive immune cell marker sets) at each of the three significance levels for the sex-DE input genes.

To further investigate patterns of sexually dimorphic expression within each gene set of interest without applying arbitrary thresholds to define sex-DE genes, we also assessed the distribution of the genes in each set into two mutually exclusive categories: higher expression in males (positive  $\log_2$ (fold difference), at any *P* value); or higher expression in females (negative  $\log_2$ (fold difference), at any *P* value). We then applied a two-sided binomial test against the distribution of male-higher and female-higher genes in the corresponding background gene set (Supplementary Data 3, 6 and 7). *P* values from all binomial tests within each data set were Bonferroni-corrected for the 31 gene sets tested.

## References

- American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders* 5th edn (American Psychiatric Association, 2013).
- Wingate, M. *et al.* Prevalence of autism spectrum disorder among children aged 8 years—Autism and Developmental Disabilities Monitoring Network, 11 sites, United States, 2010. *MMWR Surveill. Summ.* **63**, 1–21 (2014).
- Hallmayer, J. *et al.* Genetic heritability and shared environmental factors among twin pairs with autism. *Arch. Gen. Psychiatry* **68**, 1095–1102 (2011).
- Constantino, J. N., Zhang, Y., Frazier, T., Abbacchi, A. M. & Law, P. Sibling recurrence and the genetic epidemiology of autism. *Am. J. Psychiatry* **167**, 1349–1356 (2010).
- Ozonoff, S. *et al.* Recurrence risk for autism spectrum disorders: a Baby Siblings Research Consortium study. *Pediatrics* **128**, e488–e495 (2011).
- Abrahams, B. S. & Geschwind, D. H. Advances in autism genetics: on the threshold of a new neurobiology. *Nat. Rev. Genet.* **9**, 341–355 (2008).
- De Rubeis, S. *et al.* Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* **515**, 209–215 (2014).
- Iossifov, I. *et al.* The contribution of de novo coding mutations to autism spectrum disorder. *Nature* **515**, 216–221 (2014).
- Sanders, S. J. *et al.* Insights into autism spectrum disorder genomic architecture and biology from 71 risk loci. *Neuron* **87**, 1215–1233 (2015).
- Sanders, S. J. *et al.* De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* **485**, 237–241 (2012).
- Fombonne, E. Epidemiology of pervasive developmental disorders. *Pediatr. Res.* **65**, 591–598 (2009).
- Werling, D. M. & Geschwind, D. H. Sex differences in autism spectrum disorders. *Curr. Opin. Neurol.* **26**, 146–153 (2013).
- Zhao, X. *et al.* A unified genetic theory for sporadic and inherited autism. *Proc. Natl Acad. Sci. USA* **104**, 12831–12836 (2007).
- Robinson, E. B., Lichtenstein, P., Anckarsater, H., Happe, F. & Ronald, A. Examining and interpreting the female protective effect against autistic behavior. *Proc. Natl Acad. Sci. USA* **110**, 5258–5262 (2013).
- Jacquemont, S. *et al.* A higher mutational burden in females supports a 'female protective model' in neurodevelopmental disorders. *Am. J. Hum. Genet.* **94**, 415–425 (2014).
- Bailey, Jr D. B. *et al.* Autistic behavior in young boys with fragile X syndrome. *J. Autism Dev. Disord.* **28**, 499–508 (1998).
- Jamain, S. *et al.* Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. *Nat. Genet.* **34**, 27–29 (2003).
- Stone, J. L. *et al.* Evidence for sex-specific risk alleles in autism spectrum disorder. *Am. J. Hum. Genet.* **75**, 1117–1123 (2004).
- Werling, D. M., Lowe, J. K., Luo, R., Cantor, R. M. & Geschwind, D. H. Replication of linkage at chromosome 20p13 and identification of suggestive sex-differential risk loci for autism spectrum disorder. *Mol. Autism* **5**, 13 (2014).
- Skuse, D. H. Imprinting, the X-chromosome, and the male brain: explaining sex differences in the liability to autism. *Pediatr. Res.* **47**, 9–16 (2000).
- Baron-Cohen, S. The extreme male brain theory of autism. *Trends Cogn. Sci.* **6**, 248–254 (2002).
- Baron-Cohen, S. *et al.* Elevated fetal steroidogenic activity in autism. *Mol. Psychiatry* **20**, 369–376 (2014).
- Gupta, S. *et al.* Transcriptome analysis reveals dysregulation of innate immune response genes and neuronal activity-dependent genes in autism. *Nat. Commun.* **5**, 5748 (2014).
- Voineagu, I. *et al.* Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* **474**, 380–384 (2011).
- Albright, A. V. & Gonzalez-Scarano, F. Microarray analysis of activated mixed glial (microglia) and monocyte-derived macrophage gene expression. *J. Neuroimmunol.* **157**, 27–38 (2004).
- Cahoy, J. D. *et al.* A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. *J. Neurosci.* **28**, 264–278 (2008).
- Zeisel, A. *et al.* Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* **347**, 1138–1142 (2015).
- BrainSpan Consortium. BrainSpan: Atlas of the Developing Human Brain. Available: <http://www.brainspan.org> (2013) [Accessed 31 October 2013].
- Kang, H. J. *et al.* Spatio-temporal transcriptome of the human brain. *Nature* **478**, 483–489 (2011).
- Parikshak, N. N. *et al.* Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell* **155**, 1008–1021 (2013).
- Wilsey, A. J. *et al.* Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell* **155**, 997–1007 (2013).
- Basu, S. N., Kollu, R. & Banerjee-Basu, S. AutDB: a gene reference resource for autism research. *Nucleic Acids Res.* **37**, D832–D836 (2009).
- Darnell, J. C. *et al.* FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell* **146**, 247–261 (2011).
- Jojić, V. *et al.* Identification of transcriptional regulators in the mouse immune system. *Nat. Immunol.* **14**, 633–643 (2013).
- Trabzuni, D. *et al.* Widespread sex differences in gene expression and splicing in the adult human brain. *Nat. Commun.* **4**, 2771 (2013).
- Vawter, M. P. *et al.* Gender-specific gene expression in post-mortem human brain: localization to sex chromosomes. *Neuropsychopharmacology* **29**, 373–384 (2004).
- Yang, X. *et al.* Tissue-specific expression and regulation of sexually dimorphic genes in mice. *Genome Res.* **16**, 995–1004 (2006).
- Sorge, R. E. *et al.* Different immune cells mediate mechanical pain hypersensitivity in male and female mice. *Nat. Neurosci.* **18**, 1081–1083 (2015).
- Oldham, M. C. *et al.* Functional organization of the transcriptome in human brain. *Nat. Neurosci.* **11**, 1271–1282 (2008).
- Windén, K. D. *et al.* The organization of the transcriptional network in specific neuronal classes. *Mol. Syst. Biol.* **5**, 291 (2009).
- Harrow, J. *et al.* GENCODE: producing a reference annotation for ENCODE. *Genome Biol.* **7**(Suppl 1): S4 1–S4 9 (2006).
- Smail, P. J., Reyes, F. I., Winter, J. S. D. & Faiman, C. The fetal hormonal environment and its effect on the morphogenesis of the genital system. In *Pediatric Andrology* Vol. 7 (eds Kogan, S. J. & Hafez, E. S. E.) 9–19 (Martinus Nijhoff, 1981).
- Niemi, M., Ikonen, M. & Hervonen, A. in *Ciba Foundation Symposium—Endocrinology of the Testis (Colloquia on Endocrinology)* Vol. 16 (eds Wolstenholme, G. E. W.) 31–55 (Wiley, 1967).
- Rabinowitz, T., Dean, D. E., Petetot, J. M. & de Courten-Myers, G. M. Gender differences in the human cerebral cortex: more neurons in males; more processes in females. *J. Child Neurol.* **14**, 98–107 (1999).
- Mong, J. A., Kurzweil, R. L., Davis, A. M., Rocca, M. S. & McCarthy, M. M. Evidence for sexual differentiation of glia in rat brain. *Horm. Behav.* **30**, 553–562 (1996).

46. Mong, J. A. & McCarthy, M. M. Steroid-induced developmental plasticity in hypothalamic astrocytes: implications for synaptic patterning. *J. Neurobiol.* **40**, 602–619 (1999).
47. Mong, J. A. & McCarthy, M. M. Ontogeny of sexually dimorphic astrocytes in the neonatal rat arcuate. *Dev. Brain Res.* **139**, 151–158 (2002).
48. Azcoitia, I., Sierra, A. & Garcia-Segura, L. M. Localization of estrogen receptor beta-immunoreactivity in astrocytes of the adult rat brain. *Glia* **26**, 260–267 (1999).
49. Sakuma, S., Tokuhara, D., Hattori, H., Matsuoka, O. & Yamano, T. Expression of estrogen receptor alpha and beta in reactive astrocytes at the male rat hippocampus after status epilepticus. *Neuropathology* **29**, 55–62 (2009).
50. Bezzi, P. *et al.* Prostaglandins stimulate calcium-dependent glutamate release in astrocytes. *Nature* **391**, 281–285 (1998).
51. Perea, G., Navarrete, M. & Araque, A. Tripartite synapses: astrocytes process and control synaptic information. *Trends Neurosci.* **32**, 421–431 (2009).
52. Haber, M., Zhou, L. & Murai, K. K. Cooperative astrocyte and dendritic spine dynamics at hippocampal excitatory synapses. *J. Neurosci.* **26**, 8881–8891 (2006).
53. Hama, H., Hara, C., Yamaguchi, K. & Miyawaki, A. PKC signaling mediates global enhancement of excitatory synaptogenesis in neurons triggered by local contact with astrocytes. *Neuron* **41**, 405–415 (2004).
54. Muller, C. M. & Best, J. Ocular dominance plasticity in adult cat visual cortex after transplantation of cultured astrocytes. *Nature* **342**, 427–430 (1989).
55. Schafer, D. P. *et al.* Microglia sculpt postnatal neural circuits in an activity and complement-dependent manner. *Neuron* **74**, 691–705 (2012).
56. McCarthy, M. M., Todd, B. J. & Amateau, S. K. Estradiol modulation of astrocytes and the establishment of sex differences in the brain. *Ann. NY Acad. Sci.* **1007**, 283–297 (2003).
57. Auerbach, B. D., Osterweil, E. K. & Bear, M. F. Mutations causing syndromic autism define an axis of synaptic pathophysiology. *Nature* **480**, 63–68 (2011).
58. Gilman, S. R. *et al.* Rare de novo variants associated with autism implicate a large functional network of genes involved in formation and function of synapses. *Neuron* **70**, 898–907 (2011).
59. Sakai, Y. *et al.* Protein interactome reveals converging molecular pathways among autism disorders. *Sci. Transl. Med.* **3**, 86ra49 (2011).
60. Allen, L. S., Hines, M., Shryne, J. E. & Gorski, R. A. Two sexually dimorphic cell groups in the human brain. *J. Neurosci.* **9**, 497–506 (1989).
61. Parikshak, N. N., Gandal, M. J. & Geschwind, D. H. Systems biology and gene networks in neurodevelopmental and neurodegenerative disorders. *Nat. Rev. Genet.* **16**, 441–458 (2015).
62. Kirschbaum, C., Kudielka, B. M., Gaab, J., Schommer, N. C. & Hellhammer, D. H. Impact of gender, menstrual cycle phase, and oral contraceptives on the activity of the hypothalamus-pituitary-adrenal axis. *Psychosom. Med.* **61**, 154–162 (1999).
63. Hansen, K. D., Irizarry, R. A. & Wu, Z. J. Removing technical variability in RNA-seq data using conditional quantile normalization. *Biostatistics* **13**, 204–216 (2012).
64. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118–127 (2007).
65. Schroeder, A. *et al.* The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Mol. Biol.* **7**, 3 (2006).
66. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009).
67. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2014).
68. Oldham, M. C., Langfelder, P. & Horvath, S. Network methods for describing sample relationships in genomic datasets: application to Huntington's disease. *BMC Syst. Biol.* **6**, 63 (2012).
69. Smyth, G. K. in *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*. (eds Gentleman, R., Carey, V., Huber, W., Irizarry, R. & Dudoit, S.) 397–420 (Springer, 2005).
70. Iossifov, I. *et al.* De novo gene disruptions in children on the autistic spectrum. *Neuron* **74**, 285–299 (2012).

## Acknowledgements

We acknowledge data resources from the BrainSpan consortium (<http://brainspan.org>; 5RC2MH089921 to Ed Lein). This work was supported by the US National Institute of Mental Health grants (5R37MH060233 and 5R01MH094714, D.H.G.), an Autism Center for Excellence network grant (9R01MH100027) and NRSA and NIMH training fellowships (F31MH093086, D.M.W.; F30MH099886 and T32MH073526, N.N.P.). We also thank members of the Geschwind laboratory as well as Dr Stephan Sanders for helpful discussions and suggestions.

## Author contributions

D.M.W. and D.H.G. designed the study and interpreted the results; D.M.W. performed data processing and all differential expression and over-representation analyses; N.N.P. ran the preprocessing for the adult replication data set and contributed code for over-representation analyses; D.M.W. drafted the manuscript with assistance from D.H.G. and input from N.N.P.

## Additional information

**Accession codes:** Gene expression data analysed in this study come from publicly available resources (adult BrainSpan RNA-seq data, [www.brainspan.org](http://www.brainspan.org); prenatal array data, GEO accession GSE25219; and adult replication RNA-seq data, GEO accession GSE76852). Computer code used in this study is available on request.

**Supplementary Information** accompanies this paper at <http://www.nature.com/naturecommunications>

**Competing financial interests:** D.H.G. serves on the scientific advisory board for SynapDx. The remaining authors declare no competing financial interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**How to cite this article:** Werling, D. M. *et al.* Gene expression in human brain implicates sexually dimorphic pathways in autism spectrum disorders. *Nat. Commun.* **7**:10717 doi: 10.1038/ncomms10717 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>