Widespread sex differences in gene expression and splicing in the adult human brain

Citation for published version:

Digital Object Identifier (DOI):
10.1038/ncomms3771

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Nature Communications

Publisher Rights Statement:
Copyright © 2013, Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-nd/3.0/

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.
There is strong evidence to show that men and women differ in terms of neurodevelopment, neurochemistry and susceptibility to neurodegenerative and neuropsychiatric disease. The molecular basis of these differences remains unclear. Progress in this field has been hampered by the lack of genome-wide information on sex differences in gene expression and in particular splicing in the human brain. Here we address this issue by using post-mortem adult human brain and spinal cord samples originating from 137 neuropathologically confirmed control individuals to study whole-genome gene expression and splicing in 12 CNS regions. We show that sex differences in gene expression and splicing are widespread in adult human brain, being detectable in all major brain regions and involving 2.5% of all expressed genes. We give examples of genes where sex-biased expression is both disease-relevant and likely to have functional consequences, and provide evidence suggesting that sex biases in expression may reflect sex-biased gene regulatory structures.
The issue of differences between the ‘male brain’ and ‘female brain’ has been endlessly debated in the psychological literature and popular press. Sex differences in human brain structure, neurochemistry, behaviour and susceptibility to neurodegenerative and neuropsychiatric disease have all been reported. Understanding the molecular basis of observed sex differences in structure, neurochemistry, behaviour and susceptibility to disease is of obvious importance both to basic neurobiology and neuropsychopathology. However, the limited scope and power of existing studies make it difficult to place these discussions in a molecular context. The maximum number of adults investigated within any single study is 18, with only one study exploring sex-biased alternative splicing in a genome-wide manner. Furthermore, while the existence of sex-specific genetic architectures in humans has been postulated, genome-wide analyses of genetic variants associated with variation in gene expression in one sex, but not the other, have never been conducted in human brain.

To address these limitations, we analysed data from the UK Brain Expression Consortium (UKBEC) and the North American Brain Expression Consortium (NABEC). This data set is particularly valuable because (i) it is large with post-mortem samples originating from 137 neuropathologically confirmed control individuals, (ii) up to 12 central nervous system (CNS) regions have been sampled from each individual, and (iii) transcriptome profiling was performed using the Affymetrix Human Exon 1.0 ST Array, which features 1.4 million probe sets assaying expression across each individual exon. In order to explore the possibility of sex-biased gene regulatory architectures we also conducted an expression quantitative trait loci (eQTL) analysis with the aim of finding significant interactions between sex and genotype. To maximize power, we used paired genotyping and gene-level expression data provided jointly by the North American Brain Expression Consortium and the UKBEC. In this way this study provides unequivocal evidence that sex-biased gene expression in the adult human brain is widespread in terms of both the number of genes and range of brain regions involved. We also show that in some specific cases, molecular differences are likely to have functional consequences relevant to human disease and finally that sex biases in expression may reflect sex-biased gene regulatory structures.

**Results**

**Discovery of sex-biased gene expression and splicing.** In total we analysed 1,182 post-mortem brain samples dissected from frontal cortex, occipital cortex (BA17, primary visual cortex), temporal cortex, intralobular white matter, thalamus, putamen, substantia nigra, hippocampus, hypothalamus, medulla, cerebellar cortex and spinal cord. These samples originated from 137 individuals of which 101 were male and 36 were female. The demographic details related to these samples are provided in Table 1. While we found no significant difference in brain pH (Mann–Whitney \( P \)-value = 0.643, \( N = 137 \)), cause of death (\( \chi^2 \) test \( P \)-value = 0.339, \( N = 137 \)) or age at death (Mann–Whitney \( P \)-value = 0.067, \( N = 137 \)) between samples originating from men and women, we did detect a significant difference in the post-mortem interval between the sexes (Mann–Whitney \( P \)-value = 0.037, \( N = 137 \)). However, we verified that none of the findings reported in this study could be accounted for by sex differences in any of these factors (Methods).

Sex-biased expression and splicing was investigated in each brain region separately and by averaging gene-level signals across all brain regions. This analysis demonstrated that sex-biased gene expression was widespread in terms of the numbers of genes, chromosomal and range of brain regions involved (Fig. 1a). Using a false discovery rate (FDR) of 0.01 for gene-level changes (unpaired \( t \)-test, \( N \) ranging from 13–128; Supplementary Data 3) and a more stringent FDR of 0.001 for calling alternate splicing (due to the potentially higher false positive rate; unpaired \( t \)-test, \( N \) ranging from 13–128, Supplementary Data 4), we identified 448 genes with evidence of sex-biased expression, equating to 2.6% (448/17,501) of all genes expressed in the human CNS. Over 5% of regional findings are validated using qRT-PCR or have supporting evidence in the literature (Supplementary Data 5). Over 85% of the identified genes (395/448) were detected on the basis of sex-biased splicing alone, suggesting that qualitative as opposed to quantitative differences in gene expression drive sexual dimorphism in adult brain. As 95% of genes with sex-biased splicing and 34% of genes with sex-biased gene-level expression map to autosomes, sex differences were not accounted for by the sex chromosomes alone and all autosomes were involved (Fig. 1a). Furthermore, sex-biased expression was detected in all 11 brain regions, making spinal cord the exception, though even this finding is probably explained by the small sample size available for this tissue (\( N = 13 \)). Thus, we show that in common with other species, including fruit flies, fish and rodents, sex-biased gene expression and splicing is a frequent phenomenon in adult human brain.

The biological significance of sex-biased expression. We explored the biological significance of this observation by investigating our list of genes with sexually dimorphic expression for involvement in human disease. We found a significant enrichment of disease-related genes (as defined by membership in the OMIM catalogue) within our list (Yates-corrected \( \chi^2 \) \( P \)-value = 3.58 \times 10^{-9}, \( N = 448 \)). Of the 114 genes with both evidence of sex-biased expression and an OMIM entry, 12 mapped to the sex chromosomes, with the remaining 102 genes located on autosomes. Amongst these genes, 39 were associated with a disease with a sex-biased incidence, 5 were associated with diseases directly affecting the development/maintenance of the reproductive system and 3 were implicated in breast, ovarian or prostate cancer (Supplementary Data 6). These findings suggest that sex differences in gene expression in human brain relate and may even help explain well-recognised differences between men and women in disease incidence and presentation.

**Table 1** Summary of the demographic details of the individuals sampled in this study stratified by sex.

<table>
<thead>
<tr>
<th>Sex</th>
<th>( N )</th>
<th>Age (years)</th>
<th>PMI (h)</th>
<th>( pH )</th>
<th>Modal cause of death</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>36</td>
<td>64 (20–102)</td>
<td>41 (1–96)</td>
<td>6.34 (5.44–6.58)</td>
<td>IHD (30.6%)</td>
</tr>
<tr>
<td>Male</td>
<td>101</td>
<td>57 (16–91)</td>
<td>47 (1.5–99)</td>
<td>6.34 (5.42–6.63)</td>
<td>IHD (48.5%)</td>
</tr>
</tbody>
</table>

Median values and ranges are supplied.
and using the percentage of expressed genes (as documented in Supplementary Table S1) with sex-biased expression as a measure of sexual dimorphism, we found wide variation within the human CNS (Fig. 1a) and identified the primary visual cortex as the most sexually dimorphic region (1.8%). However, we recognise that this measure of sexual dimorphism has limitations because it neither accounts for the magnitude of sex differences in expression nor does it recognise major qualitative differences (such as the expression of Y chromosome genes with no X chromosome orthologue). With regard to the former concern, it should be recognised that there was considerable variability in the magnitude of sex differences in gene level expression amongst autosomal genes (median fold change = 1.29).

NRXN3, which has been implicated in autism\(^{21}\), is another disease-relevant gene identified by this analysis, but in this case due to sex-biased splicing. In common with all neurexin genes, NRXN3 has two major isoforms, \(\alpha\)-neurexins and \(\beta\)-neurexins generated through the use of two alternate promoters\(^{22}\). Whereas \(\alpha\)-neurexins were expressed similarly in men and women, \(\beta\)-neurexin expression was significantly lower in women in the thalamus (unpaired t-test, \(P\)-value = 2.67 \times 10^{-13}, N = 107; Fig. 1c,d). Given that \(\alpha\)- and \(\beta\)-neurexins have distinct functions and assuming that sex-biased splicing is not restricted to adult life, this observation could be important in understanding the higher incidence of autism in males\(^{23,24}\). Furthermore, within NRXN3's signalling pathway are genes reported to be androgen-responsive\(^{25}\) (including \(NLGN4X\), \(GRIA1\) and \(GRIA2\)), adding weight to the idea that there may indeed be differences in neurine-neuroligin signalling between males and females. In fact, amongst genes with evidence of sex-biased splicing generally there was a significant enrichment of androgen\(^{25}\) but not oestrogen-responsive\(^{26}\) genes (androgens, \(P\)-value = 9.64 \times 10^{-12}; oestrogens, \(P\)-value = 0.97, Yates-corrected \(\chi^2\) tests, \(N\) = 395). This would suggest that sex-biased splicing is a means of regulating responses to sex hormones, predominantly androgens, though this particular finding may be influenced by the fact that the majority of women included in this study would be predicted to be post-menopausal on the basis of age\(^\text{27}\) (Methods).

Within our list of genes with sex-biased expression are some notable omissions, namely \(NLGN4X\), \(PRX\) and \(TMSB4X\). These genes are worth noting because they are all non-pseudautosomal X chromosome genes with Y-linked copies (\(NLGN4Y\), \(PRKY\) and \(TMSB4Y\)) that are expressed in human brain\(^{28,29}\). Consequently in order to have two expressed alleles in both male and female brain tissue these X chromosome genes would be expected to escape X inactivation resulting in higher expression in women as compared with men\(^{30,31}\). However, in all three cases (and in contrast to genes like \(ZFX\)) expression appeared to be similar in men and women, suggesting unequal dosage of these 'functional' orthologues between the sexes (Fig. 2a–d). For example, although \(NLGN4Y\) was expressed throughout the male CNS, there was no significant sex difference in the expression of \(NLGN4X\) either by

---

**Figure 1 | Sex-biased gene expression is widespread.** (a) Bar chart showing the percentage of expressed genes with evidence of sex-biased expression or splicing in each brain region. FCTX, frontal cortex; TCTX, temporal cortex; OCTX, occipital cortex; WHMT, white matter; HIPP, hippocampus; THAL, thalamus; HYPO, hypothalamus; PUTM, putamen; SNIG, substantia nigra; MEDU, medulla; CRBL, cerebellum; SPCO, spinal cord. The number of samples analysed for each region is provided as additional labelling on the x axis. (b) RSPO1 expression in women (red) and men (blue) in the human CNS. Error bars represent the standard error of the mean (\(N\) = 13). Hypothalamus (boxed) is the only brain region showing significant sex-biased gene expression. (c) Gene structure of NRXN3 and expression levels (y axis) plotted for each probe set (x axis) covering NRXN3 in males (blue line) and females (red line) in thalamus. All differentially expressed exons map to \(\beta\)-neurexins (boxed region) indicating sex-biased expression of these transcripts. Error bars represent the Standard error of the mean (s.e.m.). (d) Successful quantitative RT–PCR validation of \(\alpha\)-neurexin and \(\beta\)-neurexin expression in thalamus in men (blue) and women (red). The 2 – \(\Delta\)C\(_v\) values have been plotted to show relative gene expression levels. Error bars represent the s.e.m. (\(N\) = 29).

[Image of chart and graph showing gene expression data.]
array or quantitative RT–PCR measurements (Fig. 2e–f). Furthermore, the pattern of expression within the CNS differed between the orthologues, suggesting that there are subtle differences in their respective functions (Fig. 2). Given that loss-of-function mutations in NLGN4X are linked to autism\(^32\), sexually dimorphic expression of NLGN4 species might also relate to the higher incidence of autism within men\(^23,24\).

Moving beyond single genes, we looked for sexually dimorphic expression of entire gene networks. Using gene set enrichment analysis (GSEA\(^33\)) we tested 71 CNS-relevant gene sets from the KEGG\(^34,35\), Reactome\(^36\) and BioCarta pathway databases (Supplementary Data 7) in all 12 CNS regions. Significant sex-biased enrichment in at least one region (\(P\)-value < 0.05, where enrichment \(P\)-values were estimated using an empirical permutation-based procedure) was detected for 12 canonical pathways of which 7 could be considered independent (Supplementary Table S2). For example, we found enrichment of two immune-related pathways in females in white matter (Fig. 3a), suggesting that sex-specific thresholds for susceptibility to immune-mediated diseases might exist and potentially explaining the higher incidence of multiple sclerosis\(^14\), a white matter disease of immune origin in women.

### Evidence in support of sex-biased gene regulation.

This finding also raises the possibility of sex-biased gene regulatory
architectures. This idea was explored by conducting an eQTL analysis with the aim of finding significant interactions between sex and genotype. To maximize power, we used paired genotyping and gene level expression data provided jointly by the North American Brain Expression Consortium\textsuperscript{12,13} and UKBEC\textsuperscript{10,11} on 390 cerebellar and 390 frontal cortex samples (121 women and 269 men). Although no signals passed a Bonferroni correction of multiple testing ($\frac{\alpha}{N} \leq 10^{-9}$; Supplementary Data 8 and 9) there was enrichment of genes related to the immune system (Benjamini-corrected modified Fisher’s Exact $P$-value $= 0.02$), interleukin 1 receptor-like 1 being an example (IL1RL1/rs34990056, $P$-value for sex-genotype interaction $= 1.77 \times 10^{-11}$, $N = 390$; Fig. 3b). While these results are not conclusive, they are consistent with the well-recognized sex-bias in the incidence of immune-related diseases\textsuperscript{37} and could help explain the sex-bias in the incidence of CNS diseases with a known immune component, such as multiple sclerosis and Alzheimer’s disease\textsuperscript{38,39}.

**Methods**

**Collection of samples analysed by Affymetrix Exon Arrays.** CNS tissues originating from 137 control individuals was collected by the Medical Research Council (MRC) Sudden Death Brain and Tissue Bank, Edinburgh, UK\textsuperscript{40}, and the Sun Health Research Institute (SHRI) an affiliate of Sun Health Corporation, USA\textsuperscript{41}. A detailed description of the samples used in the study, tissue processing and dissection is provided in the study by Trabzuni et al.\textsuperscript{10} and in Supplementary Data 1. All samples had full informed consent for retrieval and were authorized for ethically approved scientific investigation (National Hospital for Neurology and Neurosurgery and Institute of Neurology Research Ethics Committee, 10/H0716/3).

**Processing of samples analysed by Affymetrix Exon Arrays.** Total RNA was isolated from human post-mortem brain tissues using the miRNeasy 96 kit (Qiagen). The quality of total RNA was evaluated by the 2100 Bioanalyzer (Agilent) and RNA 6000 Nano Kit (Agilent) before processing with the Ambion WT Expression Kit and Affymetrix GeneChip Whole Transcript Sense Target Labelling Assay, and hybridization to the Affymetrix Exon 1.0 ST Arrays following the manufacturers’ protocols. Hybridized arrays were scanned on an Affymetrix GeneChip Scanner 3000 7G. Further details regarding RNA isolation, quality control and processing are reported in Trabzuni et al.\textsuperscript{10}. A full list of the CEL files used in this study is provided in Supplementary Data 2.

**Analysis of Affymetrix Exon Array data.** All arrays were pre-processed using Robust Multi-array Average quantile normalisation with GC background correction (GC-RMA)\textsuperscript{42} and log2 transformation in Partek’s Genomics Suite v6.6 (Partek Incorporated, USA). We also calculated the ‘detection above background metric’ (DABG) using Affymetrix Power Tools (Affymetrix). After re-mapping the Affymetrix probe sets onto human genome build 19 (GRCh37) using NetAffx annotation file HuEx_1.0-st-v2 Probeset Annotations, Release 31, we restricted analysis to 174,228 probe sets annotated with gene names, containing at least three probes with unique hybridization and DABG $P$-values $< 0.001$ in 50% of male or female individuals. The gene-level expression was calculated for up to 17,501 genes by using the median signal of probe sets corresponding to each gene.

Using the age at death and the reported age-related probability of being premenopausal in the US Caucasian population\textsuperscript{27} we predicted that 67% of the female donors ($n = 24$) were post-menopausal, whereas 17% ($n = 6$) were likely to be premenopausal. As we were unable to detect any significant differences in gene...
expression between pre- and post-menopausal women our analysis is limited to sex differences alone.

Sex-biased expression and splicing was investigated in each brain region separately using Partek’s mixed-model ANOVA (equation 1) and alternative splice ANOVA (equation 2, Partek Genomics Suite v6.6) as described below:

\[ Y_{ijklmn} = \mu + \text{Brain Bank} + Sex + \text{Scan Date} + e_{ijklmn} \]

Where \( Y_{ijklmn} \) represents the \( i \)th observation on the \( j \)th Brain Bank \( k \)th Sex \( l \)th Scan Date \( m \)th Marker ID. \( \mu \) is the common effect for the whole experiment, \( \mu \) represents the random error present in the \( i \)th observation on the \( j \)th Brain Bank \( k \)th Sex \( l \)th Scan Date. The errors \( e_{ijklmn} \) are assumed to be normally and independently distributed with mean 0 and standard deviation \( \sigma \) for all measurements. Brain Bank and Scan Date are modelled as random effects.

\[ Y_{ijklmn} = \mu + \text{Brain Bank} + Sex + \text{Scan Date} + Marker ID + \text{Sample ID} + \text{Sex} \times \text{Marker ID} + e_{ijklmn} \]

Where \( Y_{ijklmn} \) represents the \( n \)th observation on the \( i \)th Brain Bank \( j \)th Sex \( k \)th Scan Date \( l \)th Marker ID \( m \)th Sample ID, \( \mu \) is the common effect for the whole experiment, \( e_{ijklmn} \) represents the random error present in the \( n \)th observation on the \( j \)th Brain Bank \( k \)th Sex \( l \)th Scan Date \( m \)th Marker ID \( n \)th Sample ID. The errors \( e_{ijklmn} \) are assumed to be normally and independently distributed with mean 0 and standard deviation \( \sigma \) for all measurements. Marker ID is ex-on-ex-on effect (alt-splicing independent to tissue type).

In order to reduce the likelihood of false positives only probe sets called as present in both male and female samples were analysed for evidence of sex-dependent alternative splicing by sex. In all types of analysis, the date of array hybridisation and brain bank (SHRI or MRC Sudden Death Brain Bank) were included as cofactors to eliminate batch effects as discussed in detail in Trabzuni et al. 11. All P-values were corrected for multiple comparisons using the FDR step-down method. We also investigated the value of integrating data across the different brain regions based on the idea that small yet consistent differences in gene expression may exist between male and female brain samples and while such differences might not be significant in a single brain region they might be detected when all samples are considered together. In order to test this approach we calculated the average expression detected at each gene level signal across all regions for each individual. The resulting values were tested for sex-biased expression (including scan date and brain bank as covariates).

Quantitative RT-PCR. Aliquots of total RNA previously extracted from each brain region and analysed on Exon Arrays were used for validation by quantitative RT-PCR analysis. These experiments were performed on a subset of samples (\( N = 85 \)) and analysed for the expression of genes/transcripts used for mRNA analysis. After re-mapping the annotation for probes according to ReMOAT46, we restricted the analysis to probes that uniquely hybridized, were associated with gene descriptions, were located on autosomal chromosomes and that passed Illumina Detection P-values of <0.01 in >10% of male or female individuals. We also removed all probes containing SNPs that passed in the European panel of the 1,000 Genomes Project (March 2012: Integrated Phase I haplotype release version 3, based on the 2010-11 data freeze and 2012-03-14 haplotypes) with a frequency of at least 1%. This resulted in the analysis of 13,425 transcripts in cerebellum and 13,396 transcripts in frontal cortex. The resulting expression data was adjusted for age, post-mortem interval and batch effects.

Identification of sex-biased eQTLs. A combined data set of 390 individuals (121 women and 269 men) was used to identify expression QTLs that behave differently in men and women. For each probe, an expression value that exceeded 3 s.d. from the mean was considered as an outlier and removed from analysis. The outlier detection was run in men and women separately and a total of 1.2% of the expression values was removed. The QTL analysis was run for each probe against a sex, SNP and the interaction term between sex and SNP in Matrix-EQTL. The P-value for the interaction term was used to select combination of SNPs and probes for further analysis in R (http://www.r-project.org). We treated multiple expression QTLs from a tissue as one signal if the SNPs involved were clustered with linkage disequilibrium >0.50 and report the most significant expression QTL. The P-value threshold corresponding to the Bonferroni correction of multiple testing is \( \sim 3.7 \times 10^{-12} \).

References


Acknowledgements

This work was supported by the MRC through the MRC Sudden Death Brain Bank (C.S.), a Project Grant (to J.H. and M.E.W.) and Training Fellowship (G0802462 to M.R.). D.T. was supported by the King Faisal Specialist Hospital and Research Centre, Saudi Arabia. The work performed by the North American Brain Expression Consortium was supported in part by the Intramural Research Program of the National Institute on Aging, National Institutes of Health, part of the US Department of Health and Human Services; project number ZIA AG00932-04. We are grateful to the Banner Sun Health Research Institute Brain and Body Donation Program of Sun City, Arizona for the provision of human biospecimens. The Brain and Body Donation Program is supported by the National Institute of Neurological Disorders and Stroke (U24 NS073026 National Brain and Tissue Resource for Parkinson’s Disease and Related Disorders), the National Institute on Aging (P30 AG19610 Arizona Alzheimer’s Disease Core Center), the Arizona Department of Health Services (contract 211002, Arizona Alzheimer’s Research Center), the Arizona Biomedical Research Commission (contracts 4001, 0011, 05-901 and 1001 to the Arizona Parkinson’s Disease Consortium) and the Michael J. Fox Foundation for Parkinson’s Research. Computing facilities used at King’s College London were partially supported by National Institute for Health Research (NIHR) Biomedical Research Centre based at Guy’s and St Thomas’ NHS Foundation Trust and King’s College London.

Author contributions

D.T., M.R. and NABEC performed the experiments. M.R., A.R., M.E.W. and S.I. analysed the data. R.W., C.S. procured and examined the tissue. M.R., J.H. and M.W. designed the study. M.R. wrote the manuscript, which all authors edited.

Additional information

Accession codes: Affymetrix exon array data has been deposited in the NCBI Gene Expression Omnibus under accession number GSE16706. Illumina HT12–v3 Expression Beadchip array data has been deposited in the NCBI Gene Expression Omnibus under accession number GSE36192.

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

Competing financial interests: The authors declare no competing financial interests.

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