Genetic and morphological estimates of androgen exposure predict social deficits in multiple neurodevelopmental disorder cohorts.

Brooke G. McKenna[1,a], Yongchao Huang[2,a], Kévin Vervier[3], Dabney Hofamman[2], Mary Cafferata[2], Seima Al-Momani[4], Florencia Lowenthal[2], Angela Zhang[5], Jin-Young Koh[6], Savantha Thenuwara[8], Leo Brueggeman[2], Ethan Bahl[2], Tanner Koomar[2], Natalie Pottschmidt[9], Taylor Kalmus[10], Lucas Casten[2], Taylor R. Thomas[2], Jacob J. Michaelson[2,*]

1 Department of Psychology, Emory University, Atlanta, USA
2 Department of Psychiatry, University of Iowa, Iowa City, USA
3 Host-Microbiota Interactions Laboratory, Wellcome Sanger Institute, Hinxton, UK
4 Department of Psychology, University of Nebraska, Omaha, USA
5 University of Washington School of Public Health, Seattle, USA
7 Molecular Otolaryngology and Renal Research Laboratories, University of Iowa, Iowa City, USA
8 Biomedical Science Department, Iowa State University, Ames, USA
9 Department of Psychology, Pennsylvania State University, State College, USA
10 Department of Psychology, University of Washington, Seattle, USA

*** jacob-michaelson@uiowa.edu**
[a] equal contribution

Abstract

Neurodevelopmental disorders (NDDs) such as autism spectrum disorder (ASD) display a strong male bias. Androgen exposure is one paradigm for investigating this male bias, and previous work has sought to connect morphological proxies of androgen exposure, including digit ratio and facial morphology, to neurodevelopmental outcomes. The results of these studies have been inconsistent and the relationships between androgen exposure and behavior remains unclear. Here, we measured both digit ratio masculinity (DRM) and facial landmark masculinity (FLM) in the same neurodevelopmental cohort (N=763) and compared these proxies of androgen exposure to clinical and parent-reported features. We found that FLM was significantly associated with diagnostic burden in males and females ($Z = 3.1$, $p = 0.002$), while DRM was not ($Z = -1.6$, $p = 0.11$). When testing for association with parent-reported problems, we found that both FLM and DRM were positively associated with concerns about social behavior ($Z = 3.1$, $p = 0.002$; $Z = 2.1$, $p = 0.03$, respectively), also in a sex-invariant manner. Furthermore, we found evidence via polygenic risk scores (PRS) that DRM indexes masculinity via testosterone levels ($t = 2.0$, $p = 0.04$), while FLM indexes masculinity through a negative relationship with sex hormone binding globulin (SHBG) levels ($t = -2.3$, $p = 0.02$). Finally, using the SPARK cohort (N=9,419) we replicated the observed relationship between polygenic estimates of testosterone, SHBG, and social functioning ($t = -2.5$, $p = 0.01$, and $t = 4.5$, $p = 6 \times 10^{-6}$ for testosterone and SHBG, respectively). Remarkably, these quantitative sex effects on social functioning were on the same order of magnitude as the effect of binary sex itself (binary male: -0.23 ± 0.05; testosterone: -0.07 ± 0.026 per SD of PRS; SHBG: 0.11 ± 0.026 per SD of PRS). These findings and their replication in the large SPARK cohort lend strong support to the hypothesis that increasing net androgen exposure diminishes capacity for social functioning in both males and females.

1 Introduction

Most neurodevelopmental disorders (NDDs) present more often in males than females [1], and autism spectrum disorder (ASD) in particular shows a striking sex bias with a 4:1 male-to-female ratio. However, despite the evidence that male sex – and masculinity more generally, as posited by the *extreme male brain theory* [2] – is a strong risk factor for ASD, the cause of this sex bias is not yet well understood. A leading theory suggests that the male sex bias may reflect the role of male sex hormones (i.e., androgens) in the etiology of ASD. Indeed, numerous studies have found that greater levels of testosterone (an androgen commonly known for its role in male reproduction, but that also contributes to neurodevelopment in both males and females) are associated with greater ASD-related traits in both sexes [3] [4] [5]. This link has been most prominently demonstrated through investigation of *fetal* testosterone exposure (FTE), which has been shown to confer risk for both diagnosed ASD [6] [7] [8] [5] and characteristic traits of ASD in non-clinical samples [9] [10]. However, research on the influence of FTE is hindered by the practical and ethical limitations of measuring the fetal environment directly.

To overcome the challenges of directly measuring FTE, morphological and genetic measures are increasingly being leveraged in order to indirectly assess androgen-related risk. Digit ratio, or the ratio in length of the second and fourth digit (i.e, 2D:4D ratio), is a sexually dimorphic feature that is lower in males than in females [11] and has been linked to FTE [12] [13]. Numerous studies indicate that individuals with ASD exhibit lower (i.e., more masculine) digit ratios than their typically-developing counterparts [14] [15] [16] [17] [18]. However, a number of recent studies have failed to replicate the association between FTE and digit ratio [19] [20] [21], indicating that digit ratio may be a useful biomarker of ASD but its biological underpinnings require further clarification. Facial masculinity has also been proposed as a potential indicator of FTE and ASD-related risk. Greater umbilical cord testosterone levels have been associated with greater adult facial masculinity in both males and females [22], and recent studies have found that facial masculinity predicts ASD traits in both females [23] [24] and males [25] with ASD. Unexpectedly, while female subjects with ASD demonstrated more masculine faces than their typically-developing peers, males with ASD exhibited *less* masculine faces than their peers. Given that these early findings do not align with the *extreme male brain theory* (which posits that testosterone would confer greater risk regardless of sex) it remains unclear whether facial masculinity can yield insight into ASD and other NDDs and, if so, how it may reflect androgen-related risk.

Running parallel to morphological proxies of androgen exposure, genetic studies of ASD can yield insight into the hypothesized role of androgens in the etiology of NDDs. While large-scale studies have so far not identified individual genomic loci that strongly implicate sex hormone pathways as a major risk factor, some smaller studies have yielded clues worthy of further investigation. For example, one study ($n = 174$) found that three genes (the estrogen receptor *ESR2*) and two steroid enzymes, *CYP11B1* and *CYP17A1*) related to the synthesis, transport, or metabolism of sex hormones were associated with ASD-related traits in non-clinical samples [26]. Another study ($n = 1{,}171$) identified three genes – that encoded an estrogen receptor (*ESR1*), steroid enzyme (*SRD5A2*), and sex-hormone binding globulin (*SHBG*) – as conferring risk for ASD-related traits in males [27], but replication in a larger sample ($n = 10{,}654$) found that only variation in *SHBG* emerged as a significant predictor [28]. Interestingly, levels of SHBG are inversely associated with levels of active (i.e., "free") testosterone. Testosterone becomes inactive when it is bound to SHBG, and studies show that low levels of SHBG are found in individuals with excessive androgen activity [29]. These studies suggest that SHBG and other hormone-binding proteins may play an important role in the etiology of ASD through regulation of testosterone exposure.

Together, these previous studies point to several important yet unanswered questions. First,

which morphological proxies of androgen exposure are most predictive of NDD diagnosis? Second, what specific behavioral and cognitive phenotypes are most related to these proxies of androgen exposure? Third, do morphological proxies such as digit ratio and facial morphology capture the same or different molecular aspects of masculinization, and how do these mechanisms relate to neurodevelopmental risk? Finally, how do quantitative effects related to androgen exposure in both sexes compare to the binary sex effect of being a Y-chromosomal male? The answers to these questions will deliver vital insight into the nature of male bias in NDDs. Toward this end, we have assembled a neurodevelopmental cohort (N=763) that includes 1) genome-wide genotypes 2) measures of digit ratio masculinity (DRM) 3) measures of facial landmark masculinity (FLM) derived by machine learning and 4) extensive clinical and parent-reported phenotypic information. This is the first study to investigate genetic factors associated with both digital and facial masculinity and their connection to neurodevelopmental outcomes. Finally, to ensure the robustness of findings generated by investigation of our discovery cohort, we have employed SPARK, a large genetic study of autism (N=9,419 in this analysis) as a means for testing the generalization of our observations.

2 Methods

2.1 Sample: devGenes

Table 1. **Demographics (devGenes)**

	Male (N=381)		Female (N=382)	
Name	N / Mean	Percent / ± SD	N / Mean	Percent / ± SD
age	21.31	17.45	30	18.15
Caucasian	264	83.3%	273	85.6%
ASD	135	35.40%	46	12.00%
ADHD	96	25.20%	52	13.60%
ID	97	25.50%	36	9.40%
epilepsy	12	3.10%	8	2.10%
language impairment	112	29.40%	44	11.50%
depression	19	5.00%	36	9.40%
bipolar disorder	4	1.00%	9	2.40%
anxiety	28	7.30%	34	8.90%
unaffected	165	43.30%	253	66.20%

devGenes is a neurodevelopmental registry drawing participants largely from the midwestern United States and Iowa in particular. Inclusion criteria are broad, with no limit on age or presence/absence of diagnosis. Participants were informed of the study through a combination of public flyers, personal letters, follow-up phone calls, and local clinicians or community leaders. Participants met with a trained member of the research team for a single 60-minute visit in individuals' homes, at independent and state-funded clinics, and in research laboratories at the University of Iowa. Study procedures were approved by the University of Iowa's Institutional Review Board, and informed consent was obtained for each participant (IRB 201505743).

2.2 Sample: SPARK

SPARK is a nationwide (U.S.) genetic study of autism funded by the Simons Foundation [30]. Genetic and self- and parent-reported demographic and phenotypic data are made available

Table 2. Demographics (SPARK)

	Male (N=6802)		Female (N=2596)	
Name	N/ Mean	Percent / ± SD	N / Mean	Percent / ± SD
age	8.07	4.09	7.95	4.32
Caucasian	4177	61.41%	1099	42.33%
ASD	5593	82.22%	1398	53.85%
ADHD	2150	31.61%	565	21.76%
ID	959	14.10%	307	11.83%
epilepsy	246	3.62%	98	3.78%
language impairment	3165	46.53%	800	30.82%
depression	301	4.43%	125	4.82%
bipolar disorder	77	1.13 %	33	1.27%
anxiety	1012	14.88%	395	15.22%
unaffected	1198	17.61%	1209	46.57%

to researchers through `https://base.sfari.org`. We used the available genome-wide common variant genotyping data (described below) as well as results from the Social Communication Questionnaire (SCQ) [31], SPARK version 4 phenotype release.

2.3 Diagnostic and parent-report data

Participants in devGenes or their legal guardians were asked to complete a detailed survey regarding the participant's demographics, developmental history, and neurodevelopmental diagnoses. After survey completion, reported medical history and diagnoses were confirmed where possible by a member of the research team through examination of participants' University of Iowa Hospital and Clinics medical records. All devGenes data were stored in a REDCap database [32] [33]. devGenes participants were classified as "affected" for analysis purposes if they reported any of the following diagnoses: ADHD, autism, intellectual disability, language impairment (including dyslexia), epilepsy, major depression, bipolar disorder, or generalized anxiety disorder. Diagnostic burden was the total number of diagnoses reported for each participant, with levels of 0, 1, 2, or 3 or more.

2.4 Masculinity measures

Hand Scans / Digit Ratio Masculinity (DRM)

Hand scans from devGenes participants were acquired using a flatbed image scanner operated by a trained research assistant. Participants were instructed to place both hands on the scanner, fingers spread and fully extended, to obtain a single scan of all ten fingers. The lengths of index fingers (i.e., second digits [2D]) and ring fingers (i.e., fourth digits [4D]) on both hands were measured manually using Image-J [34] (Fig. 1A).

Digit lengths were then corrected for rater and scanner effects with a linear model before being used to calculate left and right 2D:4D digit ratio. Using unaffected males and females, a lowess curve was fit that defined the trend, for each sex separately, of 2D:4D ratio with respect to age. To determine the 95% confidence interval of this trend, we performed bootstrap resampling 1,000 times, each time recording the lowess curve (Fig. 1B). The mean curve of these bootstrap samples was then used as the point of reference, and all individual data points for either males or females were transformed to Z-statistics using the mean curve for that sex, and point estimates of the standard deviation. Finally, we multiplied these Z-statistics by -1 so that increasingly positive values corresponded to increasing digit ratio

masculinity. These Z-statistics (N=664) were used in all subsequent analyses as "digit ratio masculinity" (DRM).

Two-Dimensional Facial Photography / Facial Landmark Masculinity (FLM)

Facial images from devGenes participants were acquired using a Nikon D3000 camera operated by a trained research assistant. Participants were instructed to look directly at the camera and maintain a neutral facial expression with closed lips (Fig. 1E).

Facial masculinity was assessed through computational analysis of the two-dimensional facial photographs. Coordinates (x, y) of 34 points were extracted using dlib [35], a computer vision library that contains algorithms for machine learning, computer vision, and image processing, including facial landmark detection. We validated this approach by comparing it with manual annotation of the 34 facial points on a subset of 99 pictures. Results indicated significant concordance overlap of coordinates between manual annotations and dlib (average Spearman correlation: 0.962 ; 95%CI [0.959,0.966]). For the present study, we used a subset of 12 coordinates (Supplement Figure 5), given the redundancy of points within very close proximity. Landmark coordinates were normalized to a [0,1] interval in both x and y directions, as facial photographs tend to have different dimensions. The Euclidean distance between landmark pairs coordinates was then computed and used to provide 66 distance features [36].

A Random Forest classifier [37] [38] was then trained to discriminate males from females using these 66 features, as well as age in years (67 features total). Stratified sampling (via the strata and sampsize arguments to the randomForest function [39] was performed so that all eight combinations of sex, age (dichotomized at 18 years), and affected status were approximately balanced at N=75 for each of the eight groups, for each tree of the Random Forest. The ntree parameter was set to 5,000 and mtry was set to 2 to reduce the greediness of the learning and increase diversity among the trees. Out-of bag estimates of misclassification showed a 38 % misclassification rate for females and a 33 % misclassification rate for males. The proportion of votes for class: male was used as the measure of masculinity. This vote proportion was transformed with a logistic function that put the midpoint of the curve at 0.4, which represents the vote proportion threshold that maximized the out-of-bag sensitivity and specificity, as determined by the F measure. These recalibrated class probabilities were then mapped onto a standard normal distribution. Next, to both observe and correct for the effects of age and sex on this masculinity estimate, we used unaffected males and females to fit a lowess curve that defined the trend, for each sex separately, of facial masculinity with respect to age. To determine the 95% confidence interval of this trend, we performed bootstrap resampling 1,000 times, each time recording the lowess curve (Fig. 1F). The mean curve of these bootstrap samples was then used as the point of reference, and all individual data points for either males or females were transformed to Z-statistics using the mean curve for that sex, and point estimates of the standard deviation. These Z-statistics (N=747) were used in all subsequent analyses as "facial landmark masculinity" (FLM).

2.5 Genotyping and imputation (devGenes)

DNA was extracted from saliva using the Autogen QuickGene-610L kit (`www.autogen.com`). The DNA quantity was assessed using the Qubit dsDNA High Sensitivity assay kit (`www.thermofisher.com`). The mean DNA concentration was 65.5 ng\µL. The DNA quality was assessed using 1% agarose gels.

SNPs were mapped to the hg19 reference of the human genome. All Quality Control (QC) steps carried out with PLINK [40] and R [41], based on QC process described in [42]. R was used for calculating standard deviations from heterozygosity and population outliers. All

other QC was carried out in PLINK. Samples and SNPs with high global missing rate were removed. This was carried out in two stages so that highly problematic SNPs, not assayed with acceptable confidence rate, and individuals do not cause systemic problems at the more stringent threshold. First, four samples and 20,307 SNPs with missing rates above 20% were removed. Second, 17 samples and 7,255 SNPs with missing rates above 5% were removed. Then, 233,882 SNPs with very low minor allele frequency (MAF), smaller than 1% in the cohort, were removed. 42,950 SNPs which grossly defy Hardy-Weinberg principle (HWE), HWE p-value smaller than 1e-10, were removed. Samples with missing rate higher than 5% on any one autosome would have been removed, but none met this criterion. Nine samples with extreme heterozygosity rates, more than 3 standard deviations from the mean rate, were removed. Six samples that were outliers on the first 10 components of multidimensional scaling, used to determine population structure, were removed. This effectively removed samples with ethnic backgrounds not captured in 1,000Genomes [43] [44] [45], or more subtle admixture. The final number of samples used from devGenes was N=239, for 295,362 SNPs.

After QC, samples were clustered based on genotype to identify population substructure of the sample. The cohort was merged with the samples from 1,000Genomes phase 3. The combined cohort was clustered into 5 groups, representing the 5 distinct super-populations found in 1,000Genomes samples. Clustering was based on the first 10 components from multi-dimensional scaling of the combined kinship matrix of the cohort and 1,000Genomes samples. The top 20 principal components of each of the 5 clusters of the cohort were calculated separately. These components were used in downstream analyses to correct for population substructure. Samples which clustered with the European 1,000Genomes samples were used for all subsequent analyses. Samples and SNPs which passed QC were imputed to the 1,000Genomes phase 3 reference using the genipe pipeline [46]. LD was calculated and files were handled with PLINK [40]. Phasing of genotypes was done with SHAPEIT [47]. Imputation was performed by IMPUTE2 [48].

2.6 Genotyping and imputation (SPARK)

Array genotyping data was generated and processed by SPARK (Freeze three, 2019-09-12), see original publication [49] for details. SNPs were mapped to the hg19 reference of the human genome. Mapping was done with liftOver, from hg38 to hg19. All QC steps carried out with PLINK [40] and R [41], based on the QC process described in [50]. R was used for calculating standard deviations from heterozygozity and population outliers.

Samples and SNPs with missing rates above 20% were removed, then samples and SNPs with missing rates above 5% were removed. This happens in two stages so that highly problematic SNPs and individuals do not cause systemic problems at the more stringent threshold. SNPs with MAF < 1% in the cohort were removed, along with SNPs with a HWE p-value < 1e-10. Sample were removed with missing rate > 5% on any one autosome. Samples with heterozygosity rates more than 3 standard deviations from the mean rate were removed. Samples with more than 3 standard deviations from the mean on any of the first 10 MDS components were removed. Also, samples with ethnic backgrounds not captured in 1,000Genomes, or more subtle admixture were removed.

After QC, samples were clustered based on genotype to be assigned to identify population substructure of the sample. The cohort was merged with the samples from the 1,000 genomes phase 3 data. The combined cohort was clustered into 5 groups, representing the 5 distinct super-populations found in 1,000Genomes. Clustering was based on the first 10 components from multi-dimensional scaling of the combined kinship matrix of the cohort and 1,000Genomes. The top 20 principal components of each of the 5 clusters of the cohort were calculated separately. These components were used in downstream analyses to correct for population substructure. Samples which clustered with the 1,000Genomes Europeans were used for all subsequent analyses. In total, 194,410 SNPs were removed (out of 616,321 SNPs),

along with 1,922 individuals removed due to QC filters.

Samples and SNPs which passed QC were imputed to the 1,000Genomes phase 3 reference using the genipe pipeline [46]. In brief, genipe performs LD calculation and file handling with PLINK [40], phasing of genotypes done with SHAPEIT [47], and imputation by IMPUTE2 [48].

2.7 Analysis of polygenic estimates

In order to investigate the underlying genetic factors associated with masculine features, we investigated the association of masculinity (digit ratio and facial masculinity, in separate models) with polygenic risk scores relevant to neurodevelopmental conditions, sex hormones, social behavior, physical traits and cognitive performance using a subset of 239 devGenes participants. A polygenic risk score (PRS) is a sum of trait-associated alleles across many genetic loci, typically weighted by effect sizes estimated from a genome-wide association study (GWAS), which is able to investigate the presence of a genetic signal in underpowered studies and can act as a biomarker for a phenotype [51]. GWAS summary statistics were downloaded from UKbiobank [52] [53] [54] [55], Psychiatric Genomics Consortium [56] [57] [58] [59] [60] [61], and Social Science Genetic Association Consortium [62] for common psychiatric conditions, cognitive performance, physical traits and social behavior. PRS for all individuals were calculated using the PRSice tool [63]. In total, 500 p-value cutoff values were evaluated between 0 to 0.5. To test PRS-masculinity associations, a linear model was fit to predict the score using each cutoff independently (z-scaled), using the top three principal components of the masculinity-residualized PRS matrix as covariates [64].

For the SPARK replication analysis, individual items from the SCQ were residualized (using a binomial generalized linear model) for age, sex, and affected status before being mapped by quantiles onto a standard normal distribution. The p value thresholds for testosterone and SHBG PRS (0.05 and 0.1, respectively) were decided on *a priori* as the p-values that led to the maximal PRS associations in the devGenes sample. Using an ordinary least squares linear model with the residualized SCQ items, one at a time, as the response variable, and both testosterone and SHBG PRS as independent variables, we assessed the significance of the PRS association with each SCQ item (Fig. 3D). Next, to more clearly illustrate the combined effect of testosterone and SHBG, we constructed three groups: (1) top decile testosterone PRS and bottom decile SHBG PRS (high testosterone, low SHBG), (2) bottom decile testosterone and top decile SHBG (low testosterone, high SHBG), and (3) the remainder being assigned to the background group.

3 Results

3.1 Facial landmark masculinity shows more pronounced sex and diagnostic differences than 2D:4D ratio

We examined the trends, across age, of both 2D:4D ratio and FLM within the entire devGenes cohort (Fig. 1). For 2D:4D ratio (Fig. 1A), we first examined the trends of unaffected (i.e., no reported diagnoses) males and females (Fig. 1B). Females showed a steady increase in 2D:4D ratio with age, while males increased until the late teen years, followed by a dip and a plateau throughout adulthood. The 95% confidence intervals (CI) for unaffected males and females overlapped substantially until young adulthood, which would suggest lower power to detect sex differences in child/adolescent samples. Comparing affected males to unaffected males (Fig. 1C), we observed that the affected trend exceeded the unaffected 95% CI from the 20s-40s. This suggests that affected adult males show more feminine 2D:4D ratios than their unaffected counterparts. Considering the 2D:4D ratio of affected females (Fig. 1D), we

observed that the trend remained within the expectation of the unaffected females. Overall this sets out expectation that when considering gross affected/unaffected status, we would likely see little difference in 2D:4D ratio (though that expectation may not hold for more specific phenotypes).

Turning to facial masculinity (Fig. 1E), we observed a more pronounced separation of the unaffected males and females, and beginning earlier (Fig. 1F). In contrast to 2D:4D ratio, we observed that affected males appeared more masculine by facial landmarks than unaffected males, with the affected trend staying outside the 95% CI until much later in life (Fig. 1G). Similarly, we observed that affected females displayed greater facial masculinity than their unaffected counterparts (Fig. 1H), again with the affected trend remaining outside the 95% CI until very late in life (where we have very few participants). It should be noted that for all subsequent analyses, we used age- and sex-corrected measures of 2D:4D ratio (digit ratio masculinity or DRM) and facial landmark masculinity (FLM). When comparing 2D:4D ratio and facial masculinity within the same neurodevelopmental cohort, facial masculinity, with its demonstration of greater masculinity in both affected males and females, would seem to be more consistent with the assertions of the *extreme male brain theory*.

3.2 FLM is associated with diagnostic burden

Having become familiar with the basic patterns of 2D:4D ratio and facial masculinity in the same neurodevelopmental cohort, we next asked whether these morphological proxies for androgen exposure were significantly and quantitatively associated with diagnostic burden. Using ordinal regression with levels of the response variable set at 0, 1, 2, or 3+ diagnoses, we tested the association for DRM and FLM independently, using sex and age as additional covariates. We found that DRM was not significantly associated with diagnostic burden ($Z = -1.60$, $p = 0.11$, Fig. 2A), while FLM was ($Z = 3.14$, $p = 0.002$, Fig. 2B). These results suggest that FLM may have more utility as a biomarker with potential clinical relevance than DRM.

3.3 DRM and FLM are associated with parent-reported concerns about social functioning

We next turned to parent-reported concerns across a variety of domains, including social deficits, restricted and repetitive behaviors, academic performance, language, sensory issues, aggression, self-harm, eating and gastrointestinal issues, and sleep. We again used ordinal regression, with parent-report items as the response variable, oriented such that larger values represented greater levels of parental concern. Using both DRM and FLM as explanatory variables, with sex and age as covariates, we tested the association with each parent-reported item (Fig. 2C). When considering the direction of effect across these parental reports, the effects of DRM and FLM were significantly positively correlated (Pearson: $r = 0.33$, $t = 2.3$, $p = 0.02$). The only item for which both DRM and FLM achieved a significant association was where parents were asked about the frequency of social interaction with individuals outside the immediate family. Increases in both DRM and FLM were significantly associated with parental reports of less frequent social interaction by the proband ($Z = 2.1$, $p = 0.03$ and $Z = 3.2$, $p = 0.002$, respectively). FLM was positively associated with a number of other parent-reported concerns, including that the proband did not have friends, showed excessive repetition of movies or music, demonstrated mannerisms that caused the parent distress, had comparative weaknesses in reading and writing, had a lower likelihood of using multiple words to communicate, were less likely to understand sarcasm and idioms, and had disrupted sleep (all $p < 0.05$, Fig. 2C). While these associations warrant further investigation, we chose to focus on social deficits because of the converging evidence between DRM and FLM. The fact that they were both included in the same model for

significance testing, and both showed a significant association in the same direction, suggests that DRM and FLM might each capture distinct aspects of masculinization that impinge on social behavior in an additive fashion. Finally, it is worth noting that no parent-reported concern was significantly negatively associated with masculinity (i.e., none were associated with greater morphological femininity).

3.4 Polygenic association of social functioning, DRM, and FLM with testosterone and SHBG

With evidence that DRM and FLM contribute additively to social deficits and through potentially different mechanisms, we next sought to improve our mechanistic and etiological insight into these traits by testing their association with a collection of polygenic risk scores (PRS) relevant to the topic of this investigation. We first examined parent-reported social interaction frequency and its associations with polygenic estimates of total serum testosterone (UK Biobank field ID 30850), sex hormone binding globulin (SHBG, UK Biobank field ID 30830), insulin-like growth factor 1 (IGF-1, UK Biobank field ID 30770), an item from the UK Biobank measuring satisfaction with friendships (UK Biobank field ID 4570, here coded such that higher values correspond to increasing dissatisfaction), ASD risk [61], ADHD risk [60], cognitive performance, and educational attainment [62] (Fig. 3A). For this analysis, we chose to observe the general pattern of association in a qualitative fashion, given the low power for formal hypothesis testing (N=239 non-missing responses for this item). We observed a positive association with total serum testosterone PRS, a negative association with SHBG PRS, and a positive association with IGF. In addition, we observed a generally positive association with the PRS for social dissatisfaction from the UK Biobank. We observed essentially no correlation with ASD risk, slightly positive association with ADHD risk, positive association for cognitive performance, and no association with educational attainment. Together, these results suggest that a combination of low SHBG, high testosterone, and high IGF may confer risk for social deficits.

We next performed a similar analysis, this time using DRM as a the trait of interest (Fig. 3B). Given the greater power for these analyses, we visually masked associations that did not reach nominal significance (i.e., $p > 0.05$, grayed out in Fig. 3B). We found that DRM showed significant associations with testosterone PRS as well as SHBG PRS (note that total serum testosterone and SHBG show a positive genetic correlation, and that in general, as total testosterone levels rise, so too do SHBG levels, presumably to compensate for the higher testosterone). We also observed a uniformly positive association between DRM and polygenic estimates of educational attainment.

Finally, we examined polygenic associations with FLM. We observed a negative association between FLM and polygenic estimates of SHBG levels, which suggests masculinization via attenuation of SHBG, which binds and inactivates circulating testosterone. The FLM-SHBG negative association and the positive DRM-testosterone association suggest parallel/additive mechanisms for achieving higher effective androgen exposure, and further suggest that DRM and FLM are complementary measures that tap different mechanisms of masculinization. Furthermore, we observed a consistent and significant positive association between FLM and polygenic estimates of friendship dissatisfaction, as reported in the UK Biobank and in agreement with the observed relationship of increasing FLM and decreased frequency of social interactions in our devGenes cohort. Together, these results indicate that increasing testosterone and decreasing SHBG (overall corresponding to more potent net androgen exposure) underlie the observed social deficits that were indexed by our morphological estimates of masculinity, DRM and FLM.

3.5 Replication in SPARK of testosterone and SHBG association with social functioning

To test the generalization of the hypothesis that emerged from our previous analyses, that is, that increasing testosterone and decreasing SHBG levels lead to deficits in social behavior, we used SPARK, a large, US-based nationwide genetic study of autism spectrum disorders. In addition to genetic data, SPARK has a host of phenotypic data relevant to ASD. Given our working hypothesis relating to social deficits, we focused on the results of the Social Communication Questionnaire (SCQ) [31], which has 40 items designed to ascertain social communication issues typically seen in ASD. Using a bivariate linear regression (polygenic estimates of testosterone and SHBG as explanatory variables), we fit a separate model for each SCQ item (Fig. 3D). For item 19, which asks whether the proband currently has a best friend or friends, we observed significant associations for both testosterone and SHBG, in the expected directions ($t = -2.5$, $p = 0.01$, and $t = 4.5$, $p = 6 \times 10^{-6}$ for testosterone and SHBG, respectively). Of all the items on the SCQ, this item is most in line with the lead associations we observed in our devGenes discovery cohort. To more clearly demonstrate the joint effect of testosterone and SHBG, and to ensure that this replication was robust against the methodology chosen, we performed a group-based analysis. After defining three groups: high net androgen exposure (individuals in the top decile of testosterone PRS and bottom decile of SHBG PRS), low net androgen exposure (individuals in the bottom decile of testosterone PRS and top decile of SHBG PRS), and background (all remaining individuals), we used these groups in a linear model to predict the age-, sex-, and diagnosis-residualized score for SCQ item 19 (has a best friend or friends). Again the effect was significant ($t = 2.62$, $p < 0.01$) with the high net androgen exposure group showing the lowest levels of endorsement of this item, and the low net androgen exposure group showing the highest levels of endorsement. Together, these results strongly support the contribution of androgen exposure to decreased capacity for social functioning.

Finally, since sex differences in social functioning in general are expected based on prior research [65], it is worth examining the relative magnitude of the effect of binary sex (i.e., male or female) on this item, compared to the quantitative sex effects (i.e., the effects of androgen exposure regardless of sex). In a logistic regression model predicting SCQ item 19 as a binary response, and including age, sex, ASD diagnosis, testosterone PRS, and SHBG PRS as explanatory variables, we found that being a binary male had a coefficient of -0.23 ± 0.05, while testosterone had a coefficient of -0.07 ± 0.026 per SD of PRS and SHBG had a coefficient of 0.11 ± 0.026 per SD of PRS. Consequently, a difference of one SD of PRS in either of SHBG and testosterone has an effect that approximates one half to one third of the effect of being a male, respectively. This is especially noteworthy because these polygenic estimates rely only on autosomal variants (i.e., sex chromosomes do not contribute), again emphasizing that these effects of androgen exposure transcend the boundaries of binary sex.

4 Discussion

In this first study to examine both digit ratio and facial masculinity in the same genetically characterized neurodevelopmental cohort, we found strong and convergent evidence that androgen exposure leads to deficits in social functioning, even while carefully correcting for binary sex. We found that facial landmark masculinity (FLM) was predictive of diagnostic burden while digit ratio masculinity (DRM) was not. In general we found that FLM showed a greater number of clinical and parent-reported associations compared to DRM, suggesting that further pursuit of biomarkers and endophenotypes based on facial morphology may prove more fruitful than digit ratio. Particularly noteworthy is that social functioning showed convergent association with both FLM and DRM. Through analysis in connection with polygenic risk scores (PRS), our findings implicate both testosterone and SHBG levels as

complementary and additive factors that impact social functioning, and our hypothesis found additional statistical support in the large SPARK cohort. Finally, we found evidence suggesting that the autosomal genetic factors that predict testosterone and SHBG levels exert effects on social functioning that are comparable in magnitude to the effect of binary (i.e., chromosomal) sex itself, adding further evidence that, at least in the context of its effect on social functioning, sex may be more accurately described as bimodal and continuous, rather than binary and discrete.

Sex hormones affect brain structure and function through multiple mechanisms and on different time scales. Human and animal research suggests that both short-term (e.g., neurotransmitter and neuropeptide) and long-term (i.e., developmental) mechanisms downstream of androgen exposure can lead to reduced capacity for social functioning. Studies in humans that have linked fetal testosterone exposure (FTE) to later reduced quality of social relationships [66] and reduced empathic descriptions of social interactions [67] underscore the long-term processes, potentially related to neural circuit formation, that may explain androgen-related social deficits. The more immediate behavioral effects of androgens might be explained by their interactions with neuropeptides oxytocin and vasopressin, which have been well-established as key mechanisms in social behavior [68] [69]. Our results build on this previous work by providing genetic evidence that the net effect of androgen exposure, that is, both the presence of testosterone and the absence of its inhibitors (such as SHBG), is a key explanatory factor in human social functioning.

Given that even short-term changes in androgen exposure can lead to changes in cognition and behavior, our findings have potential implications for the development of new treatment options. Existing therapeutics that modulate sex hormone levels, especially those already frequently used in NDD populations, such as antiepileptics [70] may hold potential for reducing social communication deficits. Since the only current FDA-approved prescription medications for ASD, risperidone and aripiprazole, do not actually target the core diagnostic criteria of deficits in social communication and restricted and repetitive behaviors, any opportunity to address these symptoms more specifically would represent a substantial advancement in the treatment of ASD.

Despite the replication of our findings, our study has several limitations that are important to acknowledge when considering the trajectory of future work. First, we did not directly measure sex hormones, but instead used well-supported upstream (polygenic estimators) and downstream (morphological features) correlates as proxies for androgen exposure. While these are convenient to measure in large samples, further testing and refinement of hypotheses connecting genetics, sex hormones, and social functioning will require direct measurement of the relevant factors. Next, because our study was not interventional, the precise structure of causal relationships between sex hormones and social functioning remains unclear. Studies of short-term testosterone exposure have demonstrated significant effects on some behaviors such as risk taking [71] but not others, such as empathizing [72]. Further interventional research across a range of exposure durations, for example in samples of patients undergoing hormone therapy, will be needed to further elucidate the mechanisms underlying the link between androgens and social functioning.

5 Acknowledgements

We would like to thank all devGenes participants for their important contributions to this research. We would also like to thank Atticus Michaelson for his permission to use his image in Figure 1. This work was supported by the National Institutes of Health [MH105527 and DC014489 to JJM]. This work was supported by a grant from the Simons Foundation (SFARI 516716, JJM). We are grateful to all of the families in SPARK, the SPARK clinical sites and SPARK staff. We appreciate obtaining access to genetic and phenotypic data for SPARK data on SFARI Base. Approved researchers can obtain the SPARK population dataset described in

this study by applying at https://base.sfari.org. This work was also supported by a grant to the University of Iowa Institution of Clinical and Translational Science, NIH and CTSA grant number: UL1TR002537.

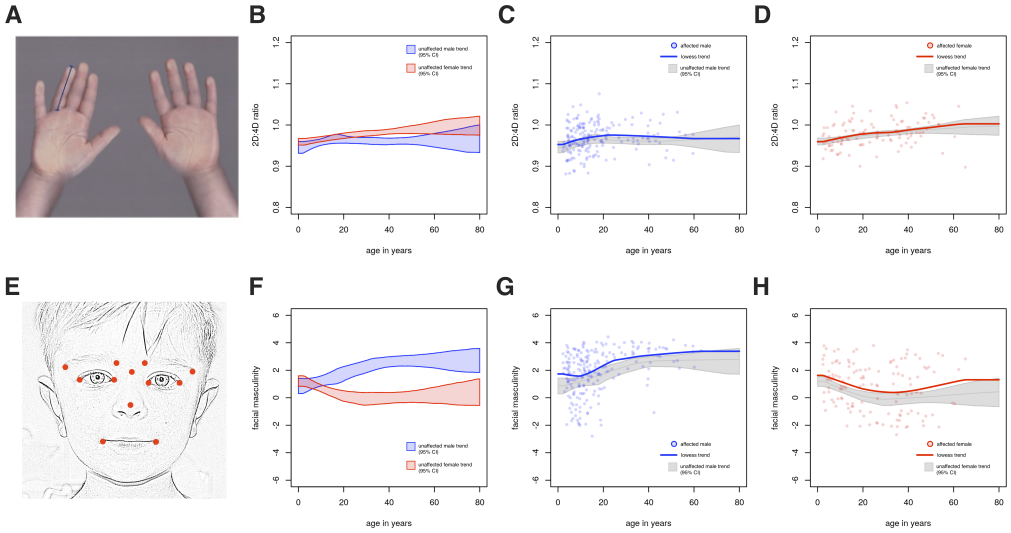


Figure 1. Masculinity as measured by digit ratio and facial landmarks. Hand scans from devGenes participants (A) were used to measure digit lengths and subsequently the ratio between the index and ring fingers (2D:4D ratio). The 95% interval of the lowess trend of 2D:4D ratio vs. age for typically developing males and females is shown (B). The 2D:4D ratio of affected males (C) and affected females (D) is shown with respect to the 95% confidence interval for their typically developing counterparts. Facial photographs from devGenes participants (E) were used to calculate facial masculinity using a Random Forest classifier, and the lowess trends and 95% confidence interval for typically developing males and females is shown (F). The facial masculinity of affected males (G) and affected females (H) is shown with respect to the 95% confidence interval for their typically developing counterparts.

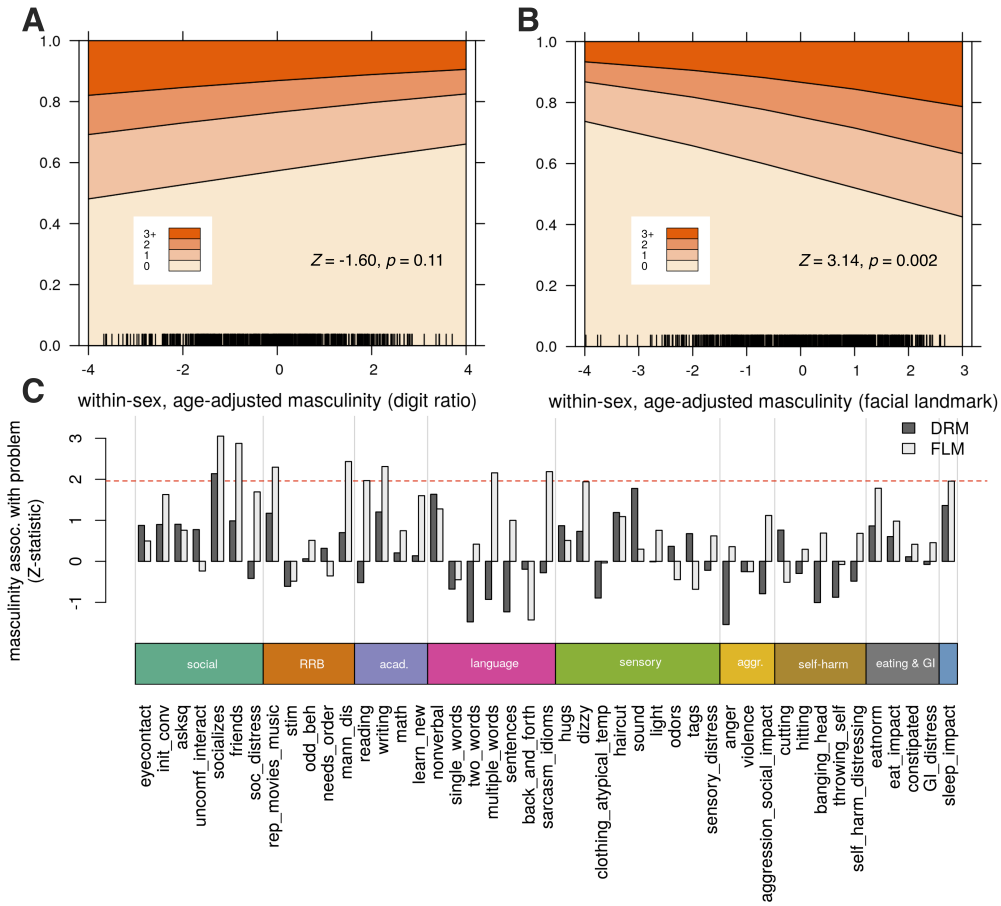


Figure 2. Masculinity associations with diagnostic burden and parent-reported problems. Ordinal regression showed no significant relationship between digit ratio masculinity (DRM) and diagnostic burden (A). In contrast, a significant positive relationship was found between facial landmark masculinity (FLM) and diagnostic burden (B). When examining parent concerns of devGenes participants across a variety of domains (C), social functioning emerged as a point of convergence, where both DRM and FLM showed positive and independent associations with parent-reported frequency of prosocial behavior.

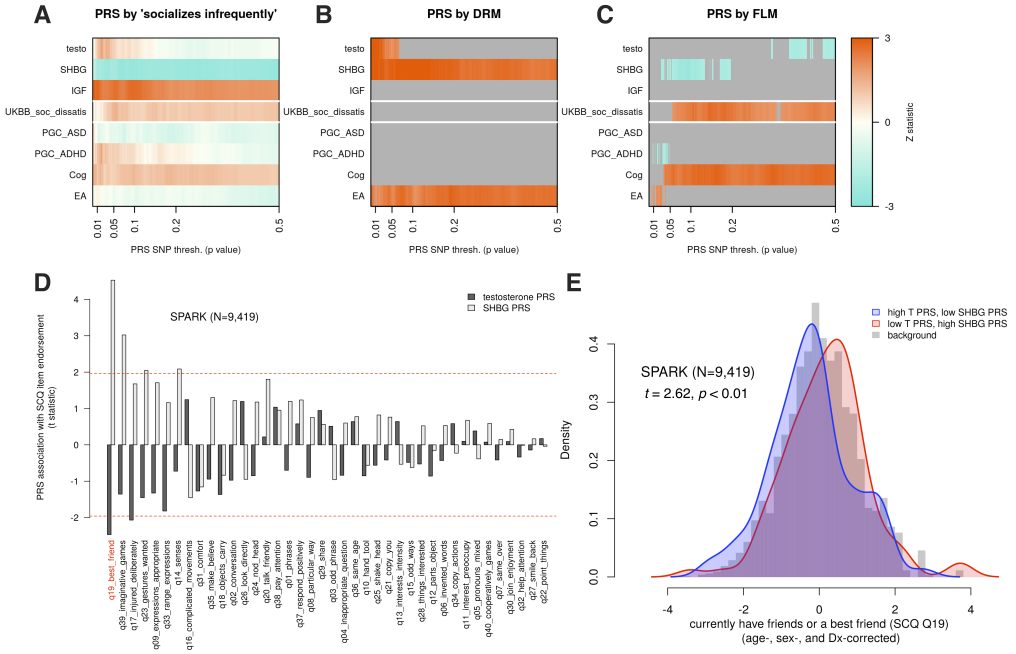


Figure 3. Polygenic associations with social functioning and morphological masculinity. Parent-reported concerns about social behavior showed distinct patterns of association with polygenic estimates related to sex hormones and social, cognitive, and neurodevelopmental traits (A). Digit ratio masculinity (DRM) showed a significant positive association with testosterone and SHBG levels, as well as educational attainment (B). Facial landmark masculinity (FLM) notably showed a negative association with polygenic SHBG levels and a positive association with polygenic dissatisfaction with friendships (C). We hypothesize that the positive DRM-testosterone association and the negative FLM-SHBG association may reflect at least part of the patterns of polygenic association seen for parent-reported concerns about social functioning. In SPARK (N=9,419), we computed polygenic testosterone and SHBG, and tested these for association with items from the SCQ, seeking to replicate our observations in panels A-C (D). A single item from the SCQ, whether the child has friends or a best friend, was significantly associated with both polygenic testosterone and SHBG in the anticipated directions. To further illustrate the combined (and opposing) effects of testosterone and SHBG on social functioning, we defined groups in the SPARK sample by being in the top (bottom) decile of polygenic testosterone (SHBG), or vice versa (E). Those who belonged to neither group were assigned to the background group. This grouping scheme shows the anticipated effect of the most potent androgen exposure (high polygenic testosterone, low polygenic SHBG) showing the lowest levels of parent-reported friendship, while those with the lowest predicted androgen exposure (low testosterone, high SHBG) showed the highest levels of parent-reported friendship.

References

1. M. Rutter, A. Caspi, and T. E. Moffitt, “Using sex differences in psychopathology to study causal mechanisms: unifying issues and research strategies,” *J Child Psychol Psychiatry,44(8)*, 2003.

2. S. Baron-Cohen, B. Auyeung, E. Ashwin, R. Knickmeyer, M. Lombardo, and B. Chakrabarti, *The Extreme Male Brain Theory of Autism: The Role of Fetal Androgens*, pp. 991–998. 05 2011.

3. R. Knickmeyer and S. Baron-Cohen, “Topical review: Fetal testosterone and sex differences in typical social development and in autism,” *Journal of Child Neurology - JCN*, vol. 21, pp. 825–845, 10 2006.

4. E. Schmidtova, S. Lakatosova, P. Celec, A. Ficek, and D. Ostatníková, “Polymorphisms in genes involved in testosterone metabolism in slovak autistic boys,” *The Endocrinologist*, vol. 20, pp. 245–249, 09 2010.

5. E. Schwarz, P. Guest, H. Rahmoune, L. Wang, Y. Levin, E. Ingudomnukul, L. Ruta, L. Kent, M. Spain, S. Baron-Cohen, and S. Bahn, “Sex-specific serum biomarker patterns in adults with asperger syndrome,” *Molecular psychiatry*, vol. 16, pp. 1213–20, 09 2010.

6. S. Baron-Cohen, D. Bowen, R. Holt, C. Allison, B. Auyeung, M. Lombardo, P. Smith, and M.-C. Lai, “The "reading the mind in the eyes" test: Complete absence of typical sex difference in 400 men and women with autism,” *PloS one*, vol. 10, p. e0136521, 08 2015.

7. D. Geier and M. Geier, “A prospective assessment of androgen levels in patients with autistic spectrum disorders: Biochemical underpinnings and suggested therapies,” *Neuro endocrinology letters*, vol. 28, pp. 565–73, 10 2007.

8. L. Ruta, E. Ingudomnukul, K. Taylor, B. Chakrabarti, and S. Baron-Cohen, “Increased serum androstenedione in adults with autism spectrum conditions,” *Psychoneuroendocrinology*, vol. 36, pp. 1154–63, 03 2011.

9. B. Auyeung, S. Baron-Cohen, E. Ashwin, R. Knickmeyer, K. Taylor, G. Hackett, and M. Hines, “Fetal testosterone predicts sexually differentiated childhood behavior in girls and in boys,” *Psychological science*, vol. 20, pp. 144–8, 02 2009.

10. S. Baron-Cohen, M. Lombardo, B. Auyeung, E. Ashwin, B. Chakrabarti, and R. Knickmeyer, “Why are autism spectrum conditions more prevalent in males?,” *PLoS biology*, vol. 9, p. e1001081, 06 2011.

11. J. Manning, D. Scutt, J. Wilson, and D. Lewis-Jones, “The ratio of 2nd to 4th digit length: A predictor of sperm numbers and concentrations of testosterone, luteinizing hormone and oestrogen,” *Human reproduction (Oxford, England)*, vol. 13, pp. 3000–4, 12 1998.

12. N. Warrington, R. Richmond, B. Fenstra, R. Myhre, R. Gaillard, L. Paternoster, C. Wang, R. Beaumont, S. Das, M. Murcia, S. Barton, A. Espinosa, E. Thiering, M. Atalay, N. Pitkänen, I. Ntalla, A. Jonsson, R. Freathy, V. Karhunen, and D. Lawlor, “Maternal and fetal genetic contribution to gestational weight gain,” *International Journal of Obesity*, vol. 42, 10 2017.

13. S. Lutchmaya, S. Baron-Cohen, P. Raggatt, R. Knickmeyer, and J. Manning, “2nd to 4th digit ratios, fetal testosterone and estradiol,” *Early human development*, vol. 77, pp. 23–8, 05 2004.

14. J. Manning, L. Barley, J. Walton, D. Lewis-Jones, R. Trivers, D. Singh, R. Thornhill, P. Rohde, T. Bereczkei, P. Henzi, M. Soler, and A. Szwed, "The 2nd:4th digit ratio, sexual dimorphism, population differences, and reproductive success: Evidence for sexually antagonistic genes?," *Evolution and human behavior : official journal of the Human Behavior and Evolution Society*, vol. 21, pp. 163–183, 06 2000.

15. E. Bruin, F. Verheij, T. Wiegman, and R. Ferdinand, "Differences in finger length ratio between males with autism, pervasive developmental disorder-not otherwise specified, adhd, and anxiety disorders," *Developmental medicine and child neurology*, vol. 48, pp. 962–5, 12 2006.

16. E. Milne, S. White, R. Campbell, J. Swettenham, P. Hansen, and F. Ramus, "Motion and form coherence detection in autistic spectrum disorder: Relationship to motor control and 2:4 digit ratio," *Journal of autism and developmental disorders*, vol. 36, pp. 225–37, 03 2006.

17. P. Noipayak, "The ratio of 2nd and 4th digit length in autistic children," *Journal of the Medical Association of Thailand = Chotmaihet thangphaet*, vol. 92, pp. 1040–5, 08 2009.

18. P. Krajmer, M. Špajdel, A. kubranská, and D. Ostatníková, "2d:4d finger ratio in slovak autism spectrum population," *Bratislavské lekárske listy*, vol. 112, pp. 377–9, 01 2011.

19. L. Lawson, J. Keelan, E. Jamnadass, M. Maybery, M. Hickey, and A. Whitehouse, "Adult digit ratio (2d:4d) is not related to umbilical cord androgen or estrogen concentrations, their ratios or net bioactivity," *Early Human Development*, vol. 91, 02 2015.

20. E. Jamnadass, J. Keelan, L. Lawson, M. Hickey, M. Maybery, and A. Whitehouse, "The perinatal androgen to estrogen ratio and autistic-like traits in the general population: A longitudinal pregnancy cohort study," *Journal of neurodevelopmental disorders*, vol. 7, p. 17, 12 2015.

21. M. Hickey, D. Doherty, R. Hart, R. Norman, E. Mattes, H. Atkinson, and D. Sloboda, "Maternal and umbilical cord androgen concentrations do not predict digit ratio (2d:4d) in girls: A prospective cohort study," *Psychoneuroendocrinology*, vol. 35, pp. 1235–44, 03 2010.

22. A. Whitehouse, S. Z. Gilani, F. Shafait, A. Mian, D. Tan, M. Maybery, J. Keelan, R. Hart, D. Handelsman, and M. Goonawardene, "Prenatal testosterone exposure is related to sexually dimorphic facial morphology in adulthood," *Proceedings of Royal Society B*, vol. 282, p. 20151351, 10 2015.

23. D. Tan, S. Russell-Smith, J. Simons, M. Maybery, D. Leung, H. Ng, and A. Whitehouse, "Perceived gender ratings for high and low scorers on the autism-spectrum quotient consistent with the extreme male brain account of autism," *PLoS ONE*, vol. 10, p. 1, 07 2015.

24. S. Bejerot and V. NORDIN, "Autismspektrumsyndrom ersätter aspergers syndrom och autism.," *Lakartidningen*, vol. 111, 01 2013.

25. S. Z. Gilani, D. Tan, S. Russell-Smith, M. Maybery, A. Mian, P. Eastwood, F. Shafait, M. Goonewardene, and A. Whitehouse, "Sexually dimorphic facial features vary according to level of autistic-like traits in the general population," *Journal of Neurodevelopmental Disorders*, vol. 7, 12 2015.

26. B. Chakrabarti, F. Dudbridge, L. Kent, S. Wheelwright, G. Hill-Cawthorne, C. Allison, S. Banerjee-Basu, and S. Baron-Cohen, "Genes related to sex steroids, neural growth, and social-emotional behavior are associated with autistic traits, empathy, and asperger syndrome," *Autism Research*, vol. 2, 06 2009.

27. A. Zettergren, L. Jonsson, D. Hovey, J. Melke, S. Lundström, P. H. Anckarsäter, MD, P. Lichtenstein, and L. Westberg, "Associations between polymorphisms in sex steroid related genes and autistic-like traits," *Psychoneuroendocrinology*, vol. 38, 07 2013.

28. A. Zettergren, S. Karlsson, D. Hovey, L. Jonsson, J. Melke, P. H. Anckarsäter, MD, P. Lichtenstein, S. Lundström, and L. Westberg, "Further investigations of the relation between polymorphisms in sex steroid related genes and autistic-like trait," *Psychoneuroendocrinology*, vol. 68, 02 2016.

29. C. Selby, "Sex hormone binding globulin: Origin, function and clinical significance," *Annals of clinical biochemistry*, vol. 27 (Pt 6), pp. 532–41, 12 1990.

30. S. Consortium, "Spark: A u.s. cohort of 50,000 families to accelerate autism research," *Neuron, 97(3)*, pp. 488–493, 2018.

31. M. Rutter, A. Bailey, and C. Lord, *The social communication questionnaire: Manual*. 2003.

32. P. Harris, R. Taylor, B. Minor, V. Elliott, M. Fernandez, L. O'Neal, L. McLeod, G. Delacqua, F. Delacqua, J. Brothier, and S. Duda, "The redcap consortium: Building an international community of software platform partners," *Journal of Biomedical Informatics*, vol. 95, p. 103208, 05 2019.

33. P. Harris, R. Taylor, R. Thielke, J. Payne, N. Gonzalez, and J. Conde, "A metadata-driven methodology and workflow process for providing translational research informatics support," *Journal of Biomedical Informatics*, vol. 42, pp. 337–381, 01 2009.

34. C. Rueden, J. Schindelin, M. Hiner, B. Dezonia, A. Walter, and K. Eliceiri, "Imagej2: Imagej for the next generation of scientific image data," *BMC Bioinformatics*, vol. 18, 01 2017.

35. D. King, "Dlib-ml: A machine learning toolkit," *Journal of Machine Learning Research*, vol. 10, pp. 1755–1758, 07 2009.

36. T. Obafemi-Ajayi, J. Miles, T. Takahashi, W. Qi, K. Aldridge, Z. Minqi, S.-Q. Xin, Y. He, and Y. Duan, "Facial structure analysis separates autism spectrum disorders into meaningful clinical subgroups," *Journal of autism and developmental disorders*, vol. 45, 10 2014.

37. L. Breiman, "Random forests, machine learning 45," *Journal of Clinical Microbiology*, vol. 2, pp. 199–228, 01 2001.

38. A. Liaw and M. Wiener, "Classification and regression by randomforest," *Forest*, vol. 23, 11 2001.

39. A. Liaw and M. Wiener, "Classification and regression by randomforest," *R News*, vol. 2, no. 3, pp. 18–22, 2002.

40. S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. Ferreira, D. Bender, J. Maller, P. Sklar, P. Bakker, M. Daly, and P. Sham, "Plink: A tool set for whole-genome association and population-based linkage analyses," *American journal of human genetics*, vol. 81, pp. 559–75, 10 2007.

41. I. Ross and G. Robert, "R: a language and environment for statistical computing. r foundation for statistical computing," *J.Comput. Graph. Stat.*, vol. 5, pp. 299–314, 01 2008.

42. S. Turner, L. Armstrong, Y. Bradford, C. Carlson, D. Crawford, A. Crenshaw, M. Andrade, K. Doheny, J. Haines, G. Hayes, G. Jarvik, L. Jiang, I. Kullo, R. Li, H. Ling, T. Manolio, M. Matsumoto, C. Mccarty, A. McDavid, and M. Ritchie, "Current protocols in human genetics," *Current protocols in human genetics / editorial board, Jonathan L. Haines ... [et al.]*, vol. Chapter 1, p. Unit1.19, 01 2011.

43. O. Delaneau, J. Marchini, K. Kim, and R. Sudbrak, "Integrating sequence and array data to create an improved 1000 genomes project haplotype reference panel," *Nature communications*, vol. 5, p. 3934, 06 2014.

44. J. O'Connell, D. Gurdasani, O. Delaneau, N. Pirastu, S. Ulivi, M. Cocca, M. Traglia, J. Huang, J. Huffman, I. Rudan, R. Mcquillan, R. Fraser, H. Campbell, O. Polasek, G. Asiki, K. Ekoru, C. Hayward, A. Wright, V. Vitart, and J. Marchini, "A general approach for haplotype phasing across the full spectrum of relatedness," *PLoS genetics*, vol. 10, p. e1004234, 04 2014.

45. A. Menelaou and J. Marchini, "Genotype calling and phasing using next-generation sequencing reads and a haplotype scaffold," *Bioinformatics (Oxford, England)*, vol. 29, 10 2012.

46. L.-P. Perreault, M.-A. Legault, G. Asselin, and M.-P. Dubé, "Genipe: An automated genome-wide imputation pipeline with automatic reporting and statistical tools," *Bioinformatics*, vol. 32, p. btw487, 08 2016.

47. O. Delaneau, J. Marchini, and J.-F. Zagury, "A linear complexity phasing method for thousands of genomes," *Nature methods*, vol. 9, pp. 179–81, 12 2011.

48. B. Howie, P. Donnelly, and J. Marchini, "A flexible and accurate genotype imputation method for the next generation of genome-wide association studies," *PLoS genetics*, vol. 5, p. e1000529, 07 2009.

49. P. Feliciano, , X. Zhou, I. Astrovskaya, T. N. Turner, T. Wang, L. Brueggeman, R. Barnard, A. Hsieh, L. G. Snyder, D. M. Muzny, A. Sabo, R. A. Gibbs, E. E. Eichler, B. J. O'Roak, J. J. Michaelson, N. Volfovsky, Y. Shen, and W. K. Chung, "Exome sequencing of 457 autism families recruited online provides evidence for autism risk genes," *npj Genomic Medicine*, vol. 4, Aug. 2019.

50. A. T. Marees, H. de Kluiver, S. Stringer, F. Vorspan, E. Curis, C. Marie-Claire, and E. M. Derks, "A tutorial on conducting genome-wide association studies: Quality control and statistical analysis," *International Journal of Methods in Psychiatric Research*, vol. 27, p. e1608, feb 2018.

51. K. Watanabe, E. Taskesen, A. Bochoven, and D. Posthuma, "Functional mapping and annotation of genetic associations with fuma," *Nature Communications*, vol. 8, 12 2017.

52. N. lab, "Uk biobank." `http://www.nealelab.is/uk-biobank/`, 2018.

53. T. Key, P. Appleby, G. Reeves, and A. Roddam, "Insulin-like growth factor 1 (igf1), igf binding protein 3 (igfbp3), and breast cancer risk: Pooled individual data analysis of 17 prospective studies," *The lancet oncology*, vol. 11, pp. 530–42, 06 2010.

54. R. Peila, R. Arthur, and T. Rohan, "Association of sex hormones with risk of cancers of the pancreas, kidney and brain in the uk biobank cohort study," *Cancer Epidemiology Biomarkers Prevention*, p. cebp.0246.2020, 06 2020.

55. K. Ruth, F. Day, J. Tyrrell, D. Thompson, A. Wood, A. Mahajan, R. Beaumont, L. Wittemans, S. Martin, A. Busch, A. M. Erzurumluoglu, B. Hollis, T. O'Mara, M. McCarthy, C. Langenberg, D. Easton, N. Wareham, S. Burgess, A. Murray, and J. Perry, "Using human genetics to understand the disease impacts of testosterone in men and women," *Nature Medicine*, vol. 26, pp. 1–7, 02 2020.

56. P. Sullivan, A. Agrawal, C. Bulik, O. Andreassen, A. Børglum, G. Breen, S. Cichon, H. Edenberg, S. Faraone, J. Gelernter, C. Mathews, C. Nievergelt, J. Smoller, and M. O'Donovan, "Psychiatric genomics: An update and an agenda," *The American journal of psychiatry*, vol. 175, p. appiajp201717030283, 10 2017.

57. R. Kuja-Halkola, J. Martin, Y. Lu, C. Hübel, C. Almqvist Malmros, P. Magnusson, H. Larsson, R. Adan, T. Ando, J. Baker, A. Bergen, W. Berrettini, A. Birgegård, C. Boni, V. Perica, H. Brandt, R. Burghardt, M. Cassina, and C. Cesta, "Associations between attention-deficit/hyperactivity disorder and various eating disorders: A swedish nationwide population study using multiple genetically informative approaches," 10 2019.

58. C.-D. Consortium, J. Smoller, N. Craddock, K. Kendler, P. Lee, B. Neale, J. Nurnberger, S. Ripke, S. Santangelo, and P. Sullivan, "Identification of risk loci with shared effects on five major psychiatric disorders: A genome-wide analysis," *Lancet*, vol. 381, 01 2013.

59. P. Group, "Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near odz4," *Nature Genetics*, vol. 44, pp. 1072–1072, 08 2012.

60. J. Martin, S. Lee, P. Mortensen, M. Nordentoft, N. Wray, B. Franke, S. Faraone, M. O'Donovan, A. Thapar, B. Neale, H.-C. Steinhausen, and S. Walitza, "A genetic investigation of sex bias in the prevalence of attention-deficit/hyperactivity disorder," *Biological Psychiatry*, vol. 83, 01 2018.

61. T. AnneyR@cardiff.ac.uk, "Meta-analysis of gwas of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia," 05 2017.

62. J. Lee, R. Wedow, A. Okbay, E. Kong, O. Maghzian, M. Zacher, T. Nguyen-Viet, P. Bowers, J. Sidorenko, R. Linnér, M. Fontana, T. Kundu, C. Lee, H. Li, R. Li, R. Royer, P. Timshel, R. Walters, E. Willoughby, and D. Cesarini, "Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals," *Nature Genetics*, vol. 50, 08 2018.

63. J. Euesden, C. Lewis, and P. O'Reilly, "Prsice: Polygenic risk score software," *Bioinformatics*, vol. 31, 12 2014.

64. D. Risso, J. Ngai, T. Speed, and S. Dudoit, "Normalization of rna-seq data using factor analysis of control genes or samples," *Nature biotechnology*, vol. 32, 08 2014.

65. P. Harris, R. Taylor, B. Minor, V. Elliott, M. Fernandez, L. O'Neal, L. McLeod, G. Delacqua, F. Delacqua, J. Brothier, and S. Duda, "The redcap consortium: Building an international community of software platform partners," *Journal of Biomedical Informatics*, vol. 95, p. 103208, 05 2019.

66. R. Knickmeyer, S. Baron-Cohen, P. Raggatt, and K. Taylor, "Foetal testosterone, social relationships, and restricted interests in children," *Journal of child psychology and psychiatry, and allied disciplines*, vol. 46, pp. 198–210, 03 2005.

67. R. Knickmeyer, S. Baron-Cohen, P. Raggatt, K. Taylor, and G. Hackett, "Fetal testosterone and empathy," *Hormones and behavior*, vol. 49, pp. 282–92, 04 2006.

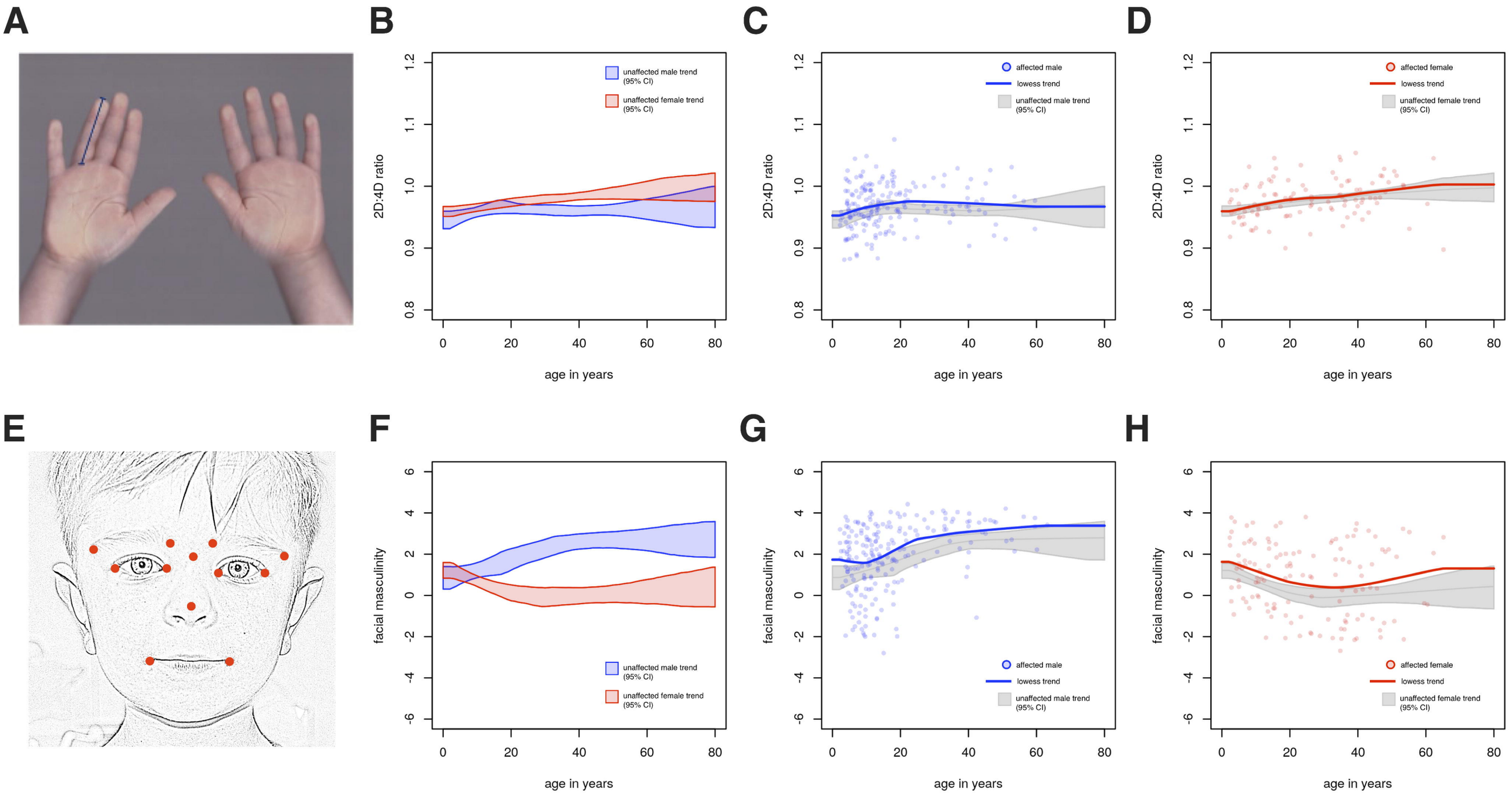
68. Q. Lu, J. Lai, Y. Du, T. Huang, P. Prukpitikul, Y. Xu, and S. Hu, “Sexual dimorphism of oxytocin and vasopressin in social cognition and behavior,” *Psychology Research and Behavior Management*, vol. Volume 12, pp. 337–349, 05 2019.

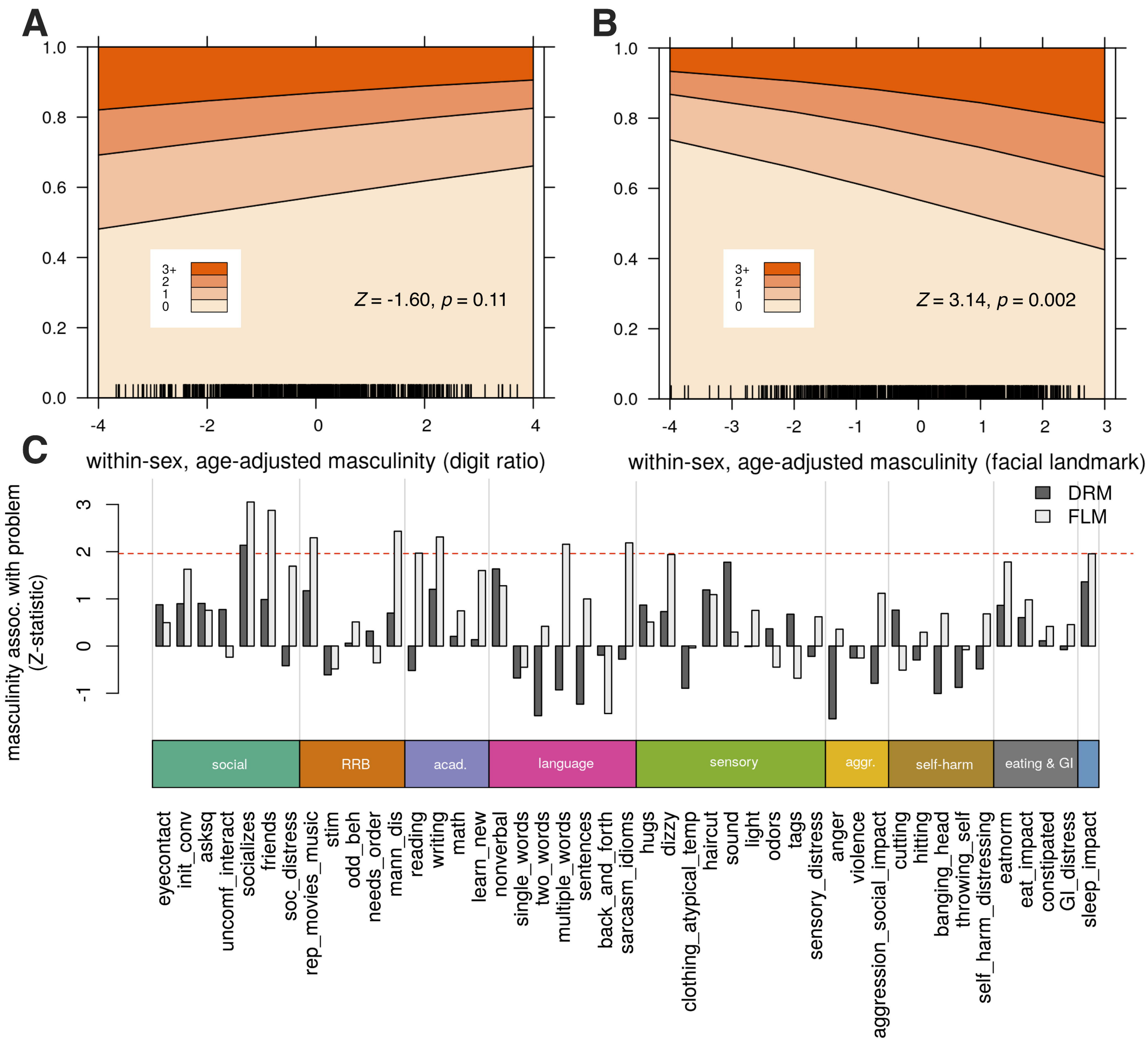
69. D. Baribeau and E. Anagnostou, “Oxytocin and vasopressin: Linking pituitary neuropeptides and their receptors to social neurocircuits,” *Frontiers in Neuroscience*, vol. 9, 09 2015.

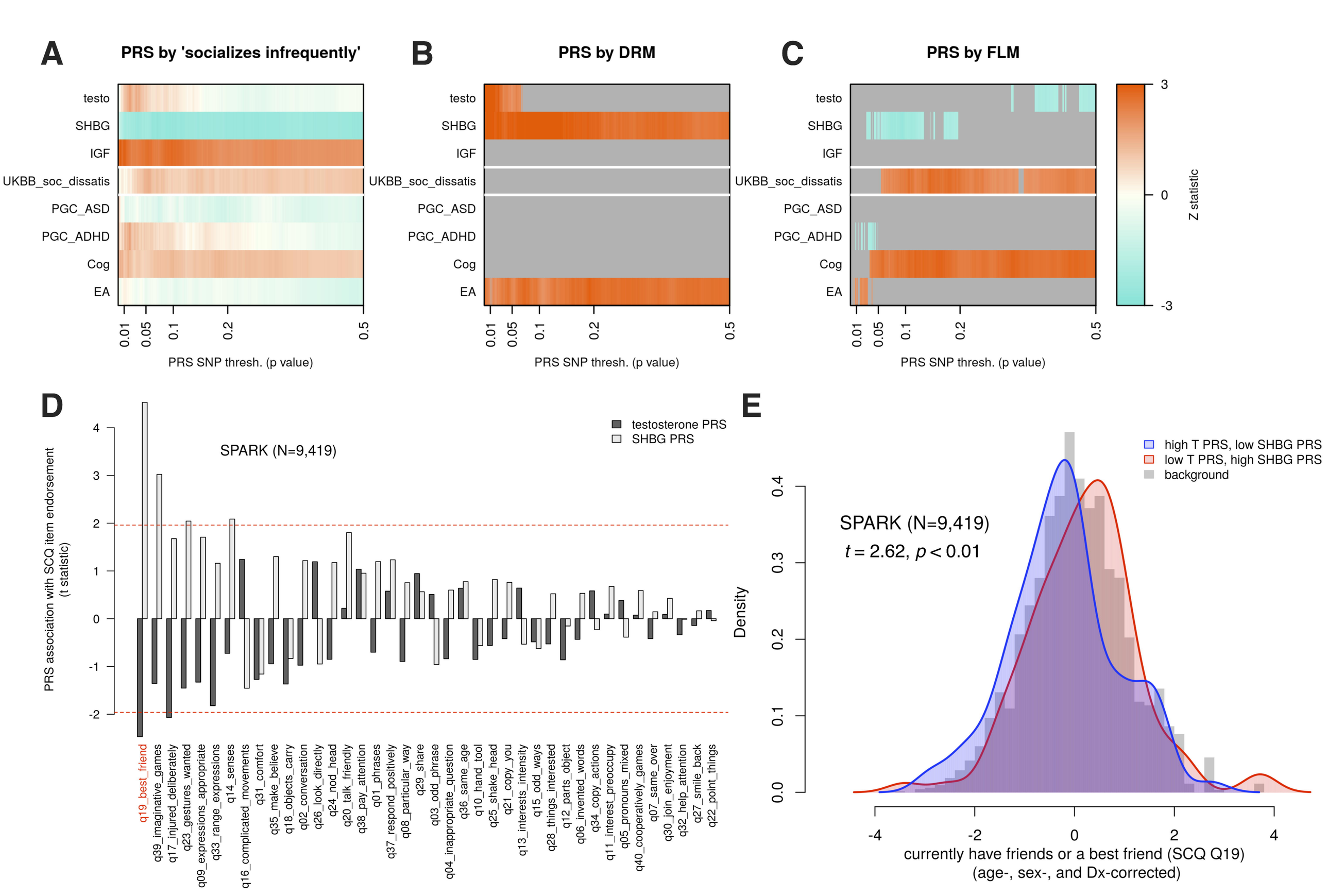
70. S. Svalheim, L. Sveberg, M. Mochol, and E. Taubøll, “Interactions between antiepileptic drugs and hormones,” *Seizure*, vol. 28, 02 2015.

71. A. Nadler, P. Jiao, C. Johnson, V. Alexander, and P. Zak, “The bull of wall street: Experimental analysis of testosterone and asset trading,” *Management Science*, vol. 64, pp. 4032–4051, 09 2018.

72. A. Nadler, D. Zava, T. Ortiz, N. Watson, J. Carré, C. Camerer, and G. Nave, “Does testosterone impair mens’ cognitive empathy? evidence from two large-scale randomized controlled trials: Supplementary material,” 01 2019.


A
B
C
D
E
F
G
H
2D:4D ratio
facial masculinity
age in years
unaffected male trend (95% CI)
unaffected female trend (95% CI)
affected male
lowess trend
affected female


A
B
C
3+
2
1
0
Z = -1.60, p = 0.11
Z = 3.14, p = 0.002
within-sex, age-adjusted masculinity (digit ratio)
within-sex, age-adjusted masculinity (facial landmark)
masculinity assoc. with problem
(Z-statistic)
DRM
FLM
social
RRB
acad.
language
sensory
aggr.
self-harm
eating & GI
eyecontact
init_conv
asksq
uncomf_interact
socializes
friends
soc_distress
rep_movies_music
stim
odd_beh
needs_order
mann_dis
reading
writing
math
learn_new
nonverbal
single_words
two_words
multiple_words
sentences
back_and_forth
sarcasm_idioms
hugs
dizzy
clothing_atypical_temp
haircut
sound
light
odors
tags
sensory_distress
anger
violence
aggression_social_impact
cutting
hitting
banging_head
throwing_self
self_harm_distressing
eatnorm
eat_impact
constipated
GI_distress
sleep_impact


A
PRS by 'socializes infrequently'
B
PRS by DRM
C
PRS by FLM
testo
SHBG
IGF
UKBB_soc_dissatis
PGC_ASD
PGC_ADHD
Cog
EA
0.01
0.05
0.1
0.2
0.5
PRS SNP thresh. (p value)
Z statistic
3
0
-3
D
SPARK (N=9,419)
testosterone PRS
SHBG PRS
PRS association with SCQ item endorsement
(t statistic)
q19_best_friend
q39_imaginative_games
q17_injured_deliberately
q23_gestures_wanted
q09_expressions_appropriate
q33_range_expressions
q14_senses
q16_complicated_movements
q31_comfort
q35_make_believe
q18_objects_carry
q02_conversation
q26_look_directly
q24_nod_head
q20_talk_friendly
q38_pay_attention
q01_phrases
q37_respond_positively
q08_particular_way
q29_share
q03_odd_phrase
q04_inappropriate_question
q36_same_age
q10_hand_tool
q25_shake_head
q21_copy_you
q13_interests_intensity
q15_odd_ways
q28_things_interested
q12_parts_object
q06_invented_words
q34_copy_actions
q11_interest_preoccupy
q05_pronouns_mixed
q40_cooperatively_games
q07_same_over
q30_join_enjoyment
q32_help_attention
q27_smile_back
q22_point_things
E
high T PRS, low SHBG PRS
low T PRS, high SHBG PRS
background
SPARK (N=9,419)
t = 2.62, p < 0.01
Density
currently have friends or a best friend (SCQ Q19)
(age-, sex-, and Dx-corrected)