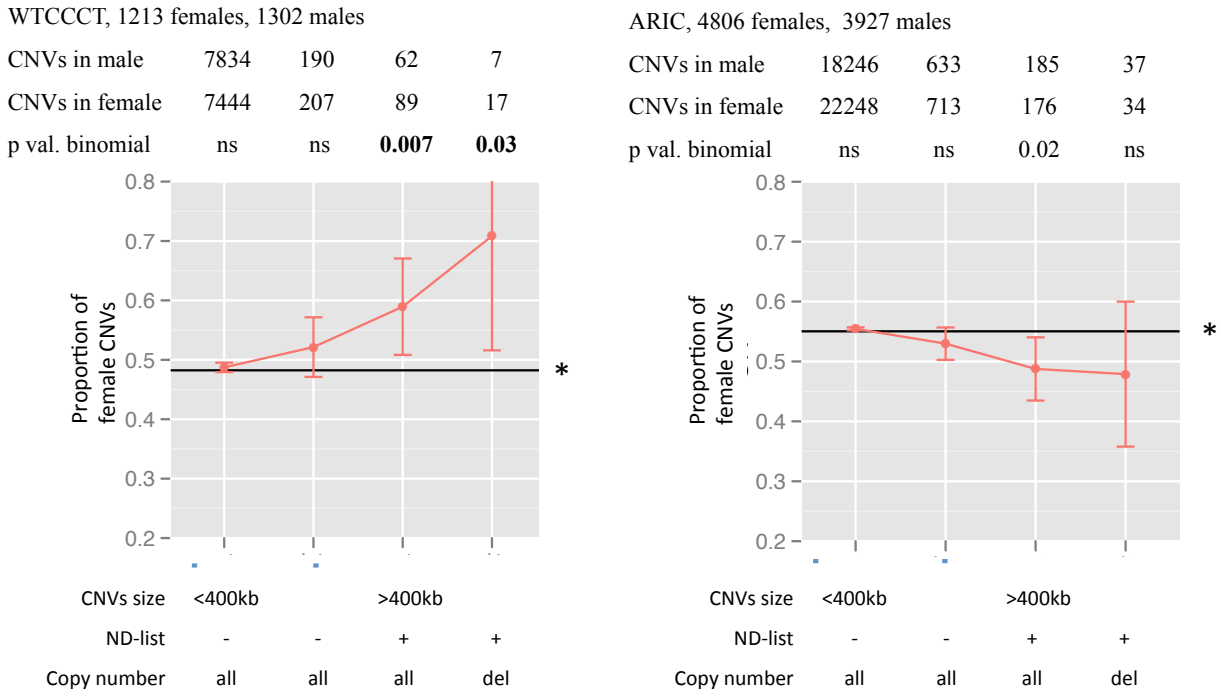
The American Journal of Human Genetics, Volume *94*

Supplemental Data

A Higher Mutational Burden in Females Supports a "Female Protective Model" in Neurodevelopmental Disorders

Sébastien Jacquemont, Bradley P. Coe, Micha Hersch, Michael H. Duyzend, Niklas Krumm, Sven Bergmann, Jacques S. Beckmann, Jill A. Rosenfeld, and Evan E. Eichler

Figure S1. Distribution of CNVs across gender in general population cohorts.



Legend : * The black horizontal line represents the proportion of females in the cohort : (0.48 in this subset of the WTCCC2 and 0.55 in ARIC). Only individuals of European descent and CNVs present in less than 1% of the data set were included in the analysis. The p-value (exact binomial test) refers to the probability that the proportion of CNVs observed in females and males is different from the gender ration of the cohort (black horizontal line). Error bars represent the 95th Confidence interval for proportion of female CNVs.

Table S1: Datasets investigated in this study

Ascertainment	proband		siblings				
Gender	males	females	males	females	Variants analyzed	Phenotypes studied	Ref
Signature Genomics (SG)	9206	6379	-	-	CNVs	NA	[1]
SSC, CNV data subset [1]	653	109			CNVs	IQ, SRS	[2]
SSC, SNV data subset [1]	226	98			SNVs	IQ, SRS	[3,4]
SG, Inheritance subset	1826*		-	-	CNVs	NA	
ISCA, Inheritance data subset	1735 *		-	-	CNVs	NA	

* All CNVs with data on inheritance were included but gender was not always available.

ISCA: International Standards for Cytogenomic Arrays

(1) The overlap between these two subsets includes 181 males and 56 females

Table S2A: Increased frequency of deleteriousness variants in females ascertained for ASD

Mutation type		**Probands** [3] 226 males, 98 females				**Unaffected siblings**[3] 70 males, 82 females			
		missense C-score >20		Truncating		Missense C-score >20		Truncating	
gene filter		all	ND or BE [2]	all	ND or BE [2]	all	ND or BE [2]	all	ND or BE [2]
All SNVs	n=	25081	3794	1787	163	11000	1645	798	80
	[1] OR	1.01	1.01	**1.1**	**1.34**	0.98	0.98	0.92	0.86
	[1]p value	ns	ns	**0.03**	**0.04**	ns	ns	ns	ns
Proband -specific SNVs	n=	20164	3066	1430	129	-	-	-	-
	[1] OR	1.02	0.99	1.09	1.30	-	-	-	-
	[1]p value	ns	ns	0.06	0.09	-	-	-	-

1: The Fisher's exact test (one-sided) tested the contingency table counting the number of nonsynonymous variants with and without the specific deleterious criteria (C-score >20, Truncating variant,) detailed in mutation type and filter (line 2 and 3 of the table).

2: ND-genes : genes involved in NDs (cf. methods). BE: genes with an expression ranked in the top 5% of all genes expressed in the brain (cf. methods). 3: Of European descent.

Table S2B: Increased deleteriousness of non-synonymous variants in females ascertained for ASD in the SSC.

	All SNVs			Missense variants			All SNVs in ND-genes			Missense variants in ND-genes		
	n=	[2]m	*p* [1]	n=	m[2]	*p* [1]	n=	m[2]	*p* [1]	n=	m[2]	*p* [1]
top 1 % *M	466	1136	**0.01**	253	24.7	ns	51	511	**0.0006**	28	24.8	ns
*F	197	1183		106	24.7		21	1484		11	24.6	
top 5 % M	2330	40	**0.002**	1267	23.5	ns	258	24	**0.001**	140	23.5	ns
F	985	43		534	23.5		107	25		57	23.6	
top 10 % M	4661	23.1	**0.003**	2535	22.8	ns	516	22.7	**0.002**	280	22.9	**0.03**
F	1970	23.3		1069	22.8		215	23.1		115	23.1	
top 20 % M	9323	21.4	**0.01**	5070	21.8	ns	1033	21.2	**0.01**	560	22	**0.03**
F	3941	21.5		2138	21.8		431	21.4		230	22.2	
top 30 % M	13984	19.8	**0.05**	7605	21	ns	1549	19.7	ns	840	21.2	**0.02**
F	5912	19.9		3207	21		647	20		345	21.4	
top 50% M	23307	14	ns	12675	18.9	ns	2583	13.7	ns	1400	19.4	ns
F	9853	14.1		5346	18.9		1079	13.7		576	19.7	

The deleteriousness of non-synonymous variants is significantly higher in females when compared to males ascertained for ASD. Truncating variants mainly drive this increased burden and the gender bias is the highest for variants involving ND-genes. Missense variants present only a marginal excess of deleteriousness in female probands. Raw C-scores of non-synonymous variants were compared between males and females. To perform the analysis on the most deleterious variants, we stratified the sample based on the top 1, 5, 10, 20, 30 and 50 percent of the C-Score distribution (cf method [5]).

[1] p values were computed by means of a Wilcoxon rank sum test. Significant p values demonstrate an increased median of C-score in females compared to males.

[2] m: Median

*M, F = values for Males and Females.

Table S3A: Increased maternal inheritance of deleterious SNVs in the SSC cohort.

	Ascertainment	Probands, n= 324				Siblings, n= 152			
	Mutation type	Non-synonymous C-score >20		Truncating		Non-synonymous C-score >20		Truncating	
	filter	all	ND/BE	all	ND/BE	all	ND/BE	all	ND/BE
SNVs	n=	24865	3756	1736	151	10943	1647	785	77
	Maternal ratio	0.50	0.51	0.51	**0.59**	0.49	0.50	0.48	0.46
	p^1 =	0.42	0.07	0.16	**0.017**	0.09	0.77	0.39	0.64
Proband specific SNVs	n=	19951	3029	1381	117	-	-	-	-
	Maternal ratio	0.50	0.51	0.52	0.59	-	-	-	-
	p^1 =	0.51	0.21	0.11	**0.032**	-	-	-	-

1: binomial test, maternal ratio is significantly higher than 50%.

C-score >20 is a deleteriousness criteria based on the Combined Annotation Dependent Depletion (CADD) [5].

ND-genes: genes involved in NDs (cf. methods). BE: genes with an expression ranked in the top 5% of all genes expressed in the brain (cf. methods).

Table S3B: Increased deleteriousness of maternally compared to paternally inherited autosomal non-synonymous variants in probands with ASD in the SSC.

	All non-synonymous variants			**Missense variants**			**All non-synonymous variants in ND-genes**			**Missense variants in ND-genes**		
	n=	**^{2}m**	**^{1}p**	**n=**	**^{2}m**	**^{1}p**	**n=**	**^{2}m**	**^{1}p**	**n=**	**^{2}m**	**^{1}p**
top 1 % Paternal	32	3479	ns	178	24	ns	35	766	ns	19	24.4	**0.03**
Maternal	32	3102		178	24		36	679		19	24.8	
top 5 % Paternal	327	1136	**0.003**	892	23	ns	179	23.7	**0.007**	96	23.4	**0.01**
Maternal	329	1168		894	23		182	24.2		98	23.7	
top 10 % Paternal	1639	37	0.05	1785	23	ns	359	22.7	**0.005**	193	22.8	0.07
Maternal	1648	42		1788	23		364	22.8		197	23	
top 20 % Paternal	3279	23	ns	3570	22	ns	719	21.1	**0.01**	386	23	ns
Maternal	3297	23		3577	22		728	21.4		394	23	
top 30 % Paternal	6556	21	ns	5355	21	ns	1079	19.5	**0.01**	579	21.2	ns
Maternal	6594	21		5365	21		1093	19.8		591	21.2	
top 50% Paternal	16391	20	ns	8926	20	ns	1799	13.3	**0.02**	965	19.3	ns
Maternal	16485	20		8943	20		1822	13.7		986	19.5	

The deleteriousness of non-synonymous variants inherited from the mother is significantly higher than those inherited from the father. This increased burden in maternally compared to paternally inherited variants is mainly driven by truncating variants involving ND-genes.

Missense variants only contribure marginaly to this difference. Raw C-scores of non-synonymous variants inherited from the mother were compared to those inherited from the father. To perform the analysis on the most deleterious variants, we stratified the sample based on the top 1, 5, 10, 20, 30 and 50 percent of the C-Score distribution.

[1] p values were computed by means of a Wilcoxon rank sum test. Significant p values demonstrate increased C-score maternally versus paternally inherited variants.

[2] m: Median. FDR correction was not applied because only 1 hypothesis was tested at several cutoffs.

Table S4: cognitive and behavioral differences between gender in different SSC subsets.

	Total SSC * (n=1961+308)			CNV subset (n=651+109)			SNV subset (n=231+102)		
	male	female	p value	male	female	P value	male	female	P value
FSIQ [1]	83	75	2×10^{-5}	85	75	5E-4	77	69	0.03
PIQ [1]	87	79	3×10^{-7}	89	76	6E-6	81	72	5×10^{-3}
VIQ [1]	80	75	0.016	82	75	5%	73	68	0.28
SRS Tscore[2]	80	90	2×10^{-17}	80	90	5E-7	82	90	7×10^{-6}

1: mean and t-test.

2: median and Wilcoxon rank sum test, due to non normal distribution.

* : IQ distribution is identical to what was reported in earlier stages of the SSC (1631 males + 256 females) [6]

Table S5: regression analyses exploring the relationship between "CNV burden" in probands, gender and global cognition.

Explained variable	Covariate	CNV burden in proband [2]	
		P-value	Effect size [unit/Mb] [3]
PIQ	-	0.001	-4.9
VIQ	-	0.02	-4.1
PIQ	gender	0.003	-4.2
VIQ	gender	0.04	-3.6
Gender*	-	0.004	-0.52
Gender*	PIQ	0.009	-0.46
Gender*	VIQ	0.005	-0.5

Two-sided regression tests

* logistic regression test

2: Sum of length of CNVs containing a "ND-gene"

3: IQ point per megabase (linear regression)

Table S6: regression analyses exploring the relationship between "SNV burden", gender and global cognition.

Explained variable	Covariate	proband SNV burden[2]	
		P-value	Effect size [unit/ratio]
PIQ	-	0.30	ns
VIQ	-	0.20	ns
PIQ	Gender	0.21	ns
VIQ	Gender	0.35	ns
Gender*	-	0.05	18.9
Gender*	PIQ	0.05	18.9
Gender*	VIQ	0.06	18.4

Two-sided regression tests
* logistic regression test
2: Ratio of total SNVs truncating ND-genes or the BE-genes over the total number of synonymous SNVs per individual.

Table S7: CNVs on the X chromosome and combined X and autosome burden analysis.

X-linked CNVs	**Male**	**Female**	**OR** [4]	**p-value** [4]
45 X	NA [1]	14	-	-
XXX / XXY [2]	24	15	0.9	0.9
>400 kb	129	129	1,45	**0.003**
>1 Mb	30	39	1.88	**0.009**
<1%, >400 kb [3], ND-list	67	82	1.8	**5×10^{-4}**
<1%, >1 Mb [3] , ND-list	27	50	2.7	**2×10^{-4}**

Estimating Autosomal + X-linked burden [5]	**Male**	**Female**	**OR** [7]	**p-value** [7]
>400 kb [3]	2764 + 136	2152 + 0 [6]	1.1	**0.003**
>1 Mb [3]	1425 + 37	1212 + 0 [6]	1.24	**6×10^{-6}**
<1 %, >400 kb, ND-list	1610 + 67	1309 + 0 [6]	1.15	**3×10^{-4}**
<1 %, >1 Mb [3], ND-list	1183 + 27	1031 + 0 [6]	1.27	**1×10^{-7}**
<0.1%, >400 kb [3], ND-list	857 + 67	743 + 0 [6]	1.2	**0.001**
<0.1%, >1 Mb [3], ND-list	576+ 27	567 + 0 [6]	1.39	**6×10^{-8}**

Legend:
CNV data from 9206 males and 6379 females ascertained for NDs (Signature Genomics).
1: The Y chromosome was not analyzed so it was not possible to count 45X males.
2: Aneuploidies (XXY and XXX) occur at an equal rate in both genders.
3: Aneuploidies (XXY and XXX) were excluded from these analyses since they occur at an equal rate in both genders and do not present with a more severe phenotype in males. Common *SHOX* and *STS* deletions were also excluded since they are not associated with neurodevelopmental symptoms.
4: Odds ratio (OR) and associated p-values represent the enrichment of X-linked CNVs in females compared to males in this group.
5: To estimate how the X-linked burden affects the results of this study, we performed the same analysis presented in Figure 1 (main text) and added the X-linked burden to the autosomal burden in males only.
6: In this very conservative approach, we considered that the X-linked burden is null in females.
7: Odds ratio (OR) and associated p-values represent the excess of mutational burden in females after accounting for the X-linked CNVs in males.
<1%, and < 0.1% relates to the frequency in the general population: rare and very rare variants.

Table S8: Comparison of behavioral and cognitive phenotypes in 1961male probands and 308 females probands from the SSC.

Cognitive / behavior measure	Male	Female	p-value
FSIQ[1]	83	75	**2×10^{-5}**
PIQ[1]	87	79	**3×10^{-7}**
VIQ[1]	80	75	**0.016**
SRS [2]	80	90	**2×10^{-17}**
VABS[1]	74	71	**4×10^{-4}**
CBCL			
Externalizing 6-18[1]	56	57.2	0.09
Internalizing 6-18[1]	60.3	59.3	0.1
ABC[1]	45.9	47.7	ns
ADOS			
Mean age at ADOS	8.9 years	8.9 years	ns
Module 1, 2 and 3 [3]	n=333, 423, 1150	n= 64, 91,144	**$p=8\times10^{-5}$, OR=1.63**
Restricted repetitive behaviors [1]	3.97	3.95	ns
Social affect [1]	9.4	9	ns
Social + communication [1]	13.2	13.7	0.07
Calibrated severity scores [1]	7.42	7.37	ns
Seizures [4]	79, 60, 42	12, 10, 9	ns

Legend:

1: Mean score and t-test.

2: Median and Wilcoxon rank sum test, due to non-normal distribution.

3: Comparison of the number of males and females who qualified for module 1, 2 or 3 of the ADOS. Males are higher functioning and there is therefore an enrichment for module 3 in males when compared to females.

4: Non-febrile seizures: 1, 2, 3+ seizures, Fisher's exact test.

Supplemental References.

1. Cooper, G.M., Coe, B.P., Girirajan, S., Rosenfeld, J.A., Vu, T.H., Baker, C., Williams, C., Stalker, H., Hamid, R., Hannig, V., et al. (2011). A copy number variation morbidity map of developmental delay. Nat. Genet. *43*, 838–846.

2. Sanders, S.J., Ercan-Sencicek, A.G., Hus, V., Luo, R., Murtha, M.T., Moreno-De-Luca, D., Chu, S.H., Moreau, M.P., Gupta, A.R., Thomson, S.A., et al. (2011). Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. Neuron *70*, 863–885.

3. O'Roak, B.J., Vives, L., Girirajan, S., Karakoc, E., Krumm, N., Coe, B.P., Levy, R., Ko, A., Lee, C., Smith, J.D., et al. (2012). Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature *485*, 246–250.

4. Sanders, S.J., Murtha, M.T., Gupta, A.R., Murdoch, J.D., Raubeson, M.J., Willsey, A.J., Ercan-Sencicek, A.G., DiLullo, N.M., Parikshak, N.N., Stein, J.L., et al. (2012). De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature *485*, 237–241.

5. Kircher, M., Witten, D.M., ORoak, B.J., Cooper, G.M., and Shendure, J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat. Genet. *in press*.

6. Fischbach, G.D., and Lord, C. (2010). The Simons Simplex Collection: A Resource for Identification of Autism Genetic Risk Factors. Neuron *68*, 192–195.