

Article

## Whole Exome Sequencing in Females with Autism Implicates Novel and Candidate Genes

Merlin G. Butler <sup>1,\*</sup>, Syed K. Rafi <sup>1,†</sup>, Waheeda Hossain <sup>1,†</sup>, Dietrich A. Stephan <sup>2,†</sup> and Ann M. Manzardo <sup>1,†</sup>

<sup>1</sup> Departments of Psychiatry & Behavioral Sciences, University of Kansas Medical Center, 3901 Rainbow Boulevard, MS 4015, Kansas City, KS 66160, USA;

E-Mails: rafigene@yahoo.com (S.K.R.); whossain@kumc.edu (W.H.); amanzardo@kumc.edu (A.M.M.)

<sup>2</sup> Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA 15260, USA;

E-Mail: dstephan@pitt.edu

† These authors contributed equally to this work.

\* Author to whom correspondence should be addressed; E-Mail: mbutler4@kumc.edu; Tel.: +1-913-588-1873; Fax: +1-913-588-1305.

Academic Editor: William Chi-shing Cho

Received: 21 November 2014 / Accepted: 31 December 2014 / Published: 7 January 2015

---

**Abstract:** Classical autism or autistic disorder belongs to a group of genetically heterogeneous conditions known as Autism Spectrum Disorders (ASD). Heritability is estimated as high as 90% for ASD with a recently reported compilation of 629 clinically relevant candidate and known genes. We chose to undertake a descriptive next generation whole exome sequencing case study of 30 well-characterized Caucasian females with autism (average age,  $7.7 \pm 2.6$  years; age range, 5 to 16 years) from multiplex families. Genomic DNA was used for whole exome sequencing via paired-end next generation sequencing approach and X chromosome inactivation status. The list of putative disease causing genes was developed from primary selection criteria using machine learning-derived classification score and other predictive parameters (GERP2, PolyPhen2, and SIFT). We narrowed the variant list to 10 to 20 genes and screened for biological significance including neural development, function and known neurological disorders. Seventy-eight genes identified met selection criteria ranging from 1 to 9 filtered variants per female. Five females presented with functional variants of X-linked genes (*ILIRAPL1*, *PIR*, *GABRQ*, *GPRASP2*, *SYTL4*) with cadherin, protocadherin and ankyrin repeat gene families

most commonly altered (e.g., *CDH6*, *FAT2*, *PCDH8*, *CTNNA3*, *ANKRD11*). Other genes related to neurogenesis and neuronal migration (e.g., *SEMA3F*, *MIDN*), were also identified.

**Keywords:** whole exome sequencing; females; autism spectrum disorder; genetic variants; X chromosome inactivation

---

## 1. Introduction

Classical autism or autistic disorder belongs to a group of genetically heterogeneous conditions known as Autism Spectrum Disorders (ASD) characterized by three clinical features: (1) marked impairments in verbal and non-verbal communication; (2) impairments in social interactions; and (3) restricted repetitive behaviors and interests [1,2]. Additional findings include lack of eye contact, tactile defensiveness and sleep disturbances with diagnosis validation at a young age requiring test instruments such as Autism Diagnostic Observation Schedule (ADOS) and Autism Diagnostic Interview-Revised (ADI-R) [3–5]. Latest estimates reveal that ASD affects about 1 of every 68 children by 8 years of age with a predominant male to female (4:1) ratio [6]. This occurrence rate is higher than that of epilepsy or Down syndrome. Congenital anomalies, intellectual disability, growth retardation and/or seizures can occur in a subset of children presenting with features related to autism. Regression of skills is observed in about 30% of cases with ASD and about 60% show intellectual disabilities [7,8]. Small or large head size is seen in about 30% of affected children with ASD with extreme macrocephaly associated with mutation of the *PTEN* tumor suppressor gene [9]. Heritability studies to identify the contribution of genetic factors in autism have shown an estimate as high as 90%. Recognized single gene conditions such as fragile X syndrome, Rett syndrome or tuberous sclerosis account for less than 20% of all cases with ASD [10].

Standard routine chromosome studies in individuals with ASD have shown abnormalities of over one dozen chromosomes. Various cytogenetic findings are reported including deletions, duplications, translocations and inversions often involving the chromosome 15q11-q13 region or the 22q11.2 band [10,11]. More advanced chromosome microarray studies are more powerful in finding cytogenetic abnormalities than routine chromosome studies. High resolution microarrays have detected recurrent small submicroscopic deletions or duplications in individuals with ASD indicating the presence of hundreds of candidate and/or known ASD genes localized to each human chromosome. In addition, there are numerous submicroscopic copy number changes, more often of the deletion type, seen in greater than 20% of patients with ASD using microarray analysis [11]. Many of these chromosome abnormalities contain genes playing a causative role in ASD. In a recent review of genetic linkage data, candidate genes and genome-wide association studies along with further advances in genetic technology including high resolution DNA microarray and next generation sequencing have led to a compilation of 629 clinically relevant candidate and known genes for ASD [12].

Given the fact that females with ASD are historically understudied, we performed whole exome sequencing of well-characterized females with classical autism from multiplex families in the search for existing or potentially new candidate genes for autism. We utilized a cohort of affected females recruited by the Autism Genetic Research Exchange (AGRE), a gene bank housing data and

biospecimens from over 2000 families ([www.AGRE.autismspeaks.org](http://www.AGRE.autismspeaks.org)). Most families had two or more affected children with autism. Identification of causative mutations (e.g., serotonin-related gene mutations and disturbed biology) could be important to guide selection of treatment options and medication use as well as to manage medical co-morbidities such as seizures, developmental regression (e.g., *MECP2* gene) or for cancer (e.g., *PTEN* gene). A cursory autism data base search revealed a large body of publications, particularly since 2008, linking autism to a wide range of genetic and environmental factors found only 3 (0.48%) clinically relevant ASD genes to be located on the Y chromosome while 68 (10.81%) clinically relevant ASD genes were recognized on the X-chromosome [12].

The preponderance of males with ASD may be attributable to the single X chromosome in males depriving the normal allelic pair of genes due to the XY sex chromosome constitution. Hence, sex chromosomes illustrate the most obvious genetic difference between men and women. All female mammals have two X chromosomes and achieve a balanced X chromosome gene expression with males by inactivating one of their X chromosomes, a process known as X chromosome inactivation (XCI) [13]. This process occurs randomly and very early in embryonic development. Once an X chromosome is “selected” for inactivation within a cell, then the same X chromosome remains inactivated in each subsequent daughter cell. Therefore, females have a mixture of cells with random expression of genes on a single X chromosome. Occasionally, XCI represents a nonrandom pattern or high skewness which is usually defined by at least 80% preferential inactivation of one of the two X chromosomes [14,15]. Skewed XCI appears to play a role in the increased incidence and presentation of diseases in females with known X-linked gene involvement such as Rett syndrome [16], X-linked intellectual disability [14], X-linked adrenoleukodystrophy [17] and possibly autism [15]. Skewed XCI may also reflect an early disruption in the developing embryo causing cell death followed by a small number of dividing cells repopulating the embryo [18–20]. XCI can be measured by using the androgen receptor (*AR*) gene located at chromosome Xq13 which contains a highly polymorphic CAG repeat region and normally inactivated on one of the X chromosomes in females. The *AR* gene is used most often to determine the XCI status [21,22]. Disorders with an imbalance in the number of males versus females affected as in autism become a likely condition to study XCI and its impact on X-linked genes in females. We suspect additional autosomal and/or sex chromosome based ASD gene(s) that either overwhelms the protective normal alleles of the second X chromosome among females, or preferentially silences them due to a plausible skewed X chromosome inactivation among neural cell lineages. Therefore, the whole exome sequence of females with ASD from multiplex families having more than one family member affected was analyzed in our study with next generation sequencing and bioinformatics with a novel machine-learning classification engine applied to identify the contributing ASD genes and the proportion of disease contributing alleles from the X chromosomes versus autosomes. The inactivation status of the two X chromosomes was determined to identify significant deviation from randomness, taking into consideration whether the X chromosome gene is known to escape inactivation, or known to act as a dominant gene.

## 2. Results

Whole exome sequencing from DNA of 30 females with autism spectrum disorders from AGRE identified between 100 and 300 genes showing genomic variants of novel or candidate genes for autism per subject with an initial classification score  $>0.5$ . To further narrow the list of putative disease causing genes, we increased the level of selection criteria by adding cutoff levels for other predictive parameters (GERP2, PolyPhen2 and SIFT) and specifically increased the classification score from  $>0.5$  to a more stringent score of  $>0.7$ . For example, the initial list of genes and genomic variants for Subject HI2898 consisted of 245 genes when using a classification score  $>0.5$  alone and then reduced to 22 genes when the cutoff levels of other predictive parameters (GERP2, PolyPhen2 and SIFT) were included at the chosen final selection criteria level. Finally, the classification score was raised to  $>0.7$  to generate the master list of 14 genes. For the 30 females with autism, the number of genes identified based on the final selection criteria ranged between 8 to 23 genes with 10 females presenting with an X-linked gene in the list of selected genes meeting the final selection criteria. The putative candidate genes and genomic variants were then subjected to further screening for biological significance and relevance for neural development, function and for causation of known neurological disorders based on published medical literature. In so doing, we identified two potentially causative genes for autism for Subject HI2898. See Table 1 for the list of genes meeting or exceeding the established final selection criteria level and screening for biological significance. Identical genomic variants were recognized for several subjects even after increasing the final level of selection criteria required to reduce the number of candidate genes and considered artifactual findings. These genes with identical genomic variants included *KCNJ12*, *MLL3*, *OR9G1* and *PCMTD1* with different mutations reported in other disease states and appeared to reflect artifacts in the present analysis of autism.

We report genomic coordinates, previously recognized SNPs (and their ID numbers), reference base to alternate base sequences (e.g., G $\rightarrow$ A), variant call format (VCF) information, SNP effects (synonymous vs non-synonymous), SNP codons, and GERP2, PolyPhen2, SIFT, Blosum, Blosum62 and the classification scores for each genomic variant identified per affected female (Table 1). Our analysis identified a total of 79 different genes containing variants (both known and novel) consisting of 12 (15%) known candidate genes for ASD, 24 (30%) paralogues of known candidate ASD genes and 43 (54%) novel genes with functional roles in nerve cell growth, adhesion and neurodevelopment or function as putative candidate ASD genes (See Table 2). The identified candidate ASD genes and paralogues were distributed across 11 (37%) and 16 (53%) of the females with autism, respectively.

Four females (HI0555, HI0890, HI1402 and HI2126) possessed a single variant of significance in clinically relevant candidate or known ASD genes considered conclusive for causality. Seventy-seven percent of subjects ( $N = 23$ ) possessed multiple variants with possible influences on causation of ASD. Two (9%) of the 23 females (7% of the total sample population) possessed more than five variants/mutations of known candidate ASD genes, paralogues or other functionally relevant genes with presumed influence over the development of ASD. Three (10%) of the total females (HI1143, HI1884 and HI2278) possessed a single gene variant with possible functional relevance toward the development of ASD but no known candidate ASD genes or paralogues identified to be considered conclusive.

**Table 1.** Putative disease causing genes for autism identified using whole exome sequencing of females with autism spectrum disorder (ASD).

ID Number	Gene Symbol (Category) *	Chromosome Position (Hg19)	Genomic Variant					Classification Score	GERP2	PolyPhen2 Score	SIFT Score	Blosum Score/Blosum62
			SNP Effect	Amino Acid	SNP Codon	SNP Function	VCF Allele Depth					
<b>HI0405</b> 46%:54%	<i>KCNC2</i> (P) [12]	12:75444895	Non-synon	p.F296C	tTt/tGt	Missense	162, 30	0.790	0	0.984	0	0.787/-2
	<i>ASPM</i> (F) [23]	1:197112823	Non-synon	p.R186G	Aga/Gga	Missense	69, 53	0.777	0.007	0.994	0	0.787/-2
	<i>TAF3</i> (F) [24]	10:8006662	Non-synon	p.R396G	Cga/Gga	Missense	69, 60	0.740	0	0.956	0	0.787/-2
	<i>FLRT2</i> (F) [25]	14:86089552	Non-synon	p.H564R	cAt/cGt	Missense	28, 29	0.736	0	0.958	0.02	0.787/0
	<i>SYTL4</i> (F) [26]	X:99941091	Non-synon	p.H448D	Cat/Gat	Missense	36, 37	0.710	0.006	1	0.03	0.787/-1
<b>HI0555/58%:42%</b>	<i>SETD2</i> (A) [12,27,28]	3:47164293	Non-synon	p.K610O	aaG/aaT	Missense	23, 21	0.778	0	0.970	0	8.522/0
<b>HI0558</b> 52%:48%	<i>IFT122</i> (F) [29]	3:129238526	Non-synon	p.R986H	cGc/cAc	Missense	58, 61	0.795	0.008	0.989	0	0.074/0
	<i>NUP98</i> (F) [30]	11:3756456	Non-synon	p.R519W	Cgg/Tgg	Missense	12, 9	0.793	0.006	0.999	0	24.449/-3
	<i>ZKSCAN1</i> (F) [31]	7:99621487	Non-synon	p.R119C	Cgc/Tgc	Missense	8, 5	0.782	0.006	1	0	0.787/-3
	<i>DNAH8</i> (F) [32]	6:38834433	Non-synon	p.M1971V	Atg/Gtg	Missense	8, 14	0.783	0.009	0.902	0	4.182/1
	<i>CENPJ</i> (F) [33]	13:25487103	Non-synon	p.M20L	Atg/Ttg	Missense	45, 42	0.761	0.007	0.991	0.01	4.312/2
	<i>TRIM5</i> (F) [34]	11:5878550	Non-synon	p.R127P	cGc/cCc	Missense	42, 24	0.742	0.0089	0.971	0	0.787/-2
	<i>PCDH8</i> (A) [35]	13:53421358	Non-synon	p.A404V	gCg/gTg	Missense	13, 14	0.731	0	0.924	0	0.787/0
	<i>CTNNA3</i> (A) [36]	10:68381521	Non-synon	p.M434V	Atg/Gtg	Missense	30, 17	0.723	0.009	0.942	0.01	4.183/1
<i>PPP1R18</i> (P) [37]	6:30652604	Non-synon	p.S397P	Tct/Cct	Missense	12, 6	0.703	0.005	0.958	0	0.787/-1	
<b>HI0605</b> 53%:47%	<i>CCDC64</i> (A) [38]	12:120428012	Non-synon	p.A113T	Gcc/Acc	Missense	13, 24	0.794	0.006	0.96	0	4.774/-1
	<i>DCHS1</i> (F) [39]	11:6662745	Codon Change Plus Codon Insertion	p.L32LW	ctg/ctCTGg	None	11, 8	0.700	0	0.979	0.02	0/0
<b>HI0714</b> 40%:60%	<i>TRAK2</i> (F) [40]	2:202264189	Non-synon	p.R130G	Cga/Gga	Missense	31, 27	0.805	0.010	1	0	0.064/-2
	<i>MPST</i> (F) [41]	22:37425405	Utr_3_Prime	p.Null- 1Null	Null	None	20, 21	0.731	0.008	1	0	0.787/-2
	<i>FSTL3</i> (F) [42]	19:680331	Non-synon	p.R115H	cGc/cAc	Missense	7, 3	0.729	0.010	1	0.01	8.723/0
<b>HI0751</b> 16%:84%	<i>ISTI</i> (F) [43,44]	16:71956522	Non-synon	p.P232R	cCc/cGc	Missense	63, 50	0.754	0.008	0.972	0.01	0.787/-2
	<i>KIF27</i> (P) [43,44]	9:86474259	Non-synon	p.S921P	Tca/Cca	Missense	22, 25	0.730	0.009	0.970	0.01	0.787/-1
	<i>PCDHGAI</i> (P) [45]	5:140870234	Non-synon	p.V475A	gTg/gCg	Missense	152, 141	0.713	0	0.995	0	0.787/0

Table 1. Cont.

ID Number	Gene Symbol (Category) *	Chromosome Position (Hg19)	Genomic Variant					Classification Score	GERP2	PolyPhen2 Score	SIFT Score	Blosum Score/Blosum62
			SNP Effect	Amino Acid	SNP Codon	SNP Function	VCF Allele Depth					
<b>HI0765</b> 77%:23%	<i>SEMA3F</i> (P) [46]	3:50225525	Non-synon	p.R679W	Cgg/Tgg	Missense	21, 7	0.793	0.006	1	0	0.787/-3
	<i>SCN11A</i> (P) [47]	3:38968409	Frame-shift	p.-166SS?	-/TCTT CACT	None	23, 17	0.748	0.008	0.982	0	0/0
	<i>GPRASP2</i> (F) [48]	X:101972234	Non-synon	p.R812C	Cgt/Tgt	Missense	69, 51	0.706	0	0.989	0.01	0.787/-3
<b>HI0793</b> 37%:64%	<i>ELTD1</i> (F) [49]	1:79356850	Non-synon	p.C687R	Tgt/Cgt	Missense	50, 25	0.807	0.008	0.998	0	10.199/-3
	<i>DENND3</i> (F) [50]	8:142161854	Non-synon	p.S252L	tCg/tTg	Missense	40, 20	0.806	0.006	0.992	0	0.787/-2
	<i>MIDN</i> (F) [51]	19:1250353	Non-synon	p.C19G	Tgc/Ggc	Missense	8, 3	0.793	0.008	0.998	0	17.850/-3
	<i>RADIL</i> (F) [52]	7:4917586	Non-synon	p.P61R	cCt/cGt	Missense	42, 27	0.774	0.004	1	0	14.302/-2
	<i>EXOC7</i> (P) [53,54]	17:74097856	Non-synon	p.R71Q	cGg/cAg	Missense	36, 46	0.755	0.006	0.992	0.01	0.059/1
	<i>PLCE1</i> (F) [55]	10:96064250	Non-synon	p.H1515Y	Cat/Tat	Missense	22, 14	0.737	0.008	0.999	0.02	7.078/2
	<i>KDM5A</i> (F) [56]	12:416817	Non-synon	p.A1244T	Gcc/Acc	Missense	42, 25	0.706	0	0.993	0	0.157/0
<i>ARAP2</i> (F) [57]	4:36115873	Non-synon	p.D1358N	Gat/Aat	Missense	70, 53	0.702	0.008	0.907	0.02	12.052/1	
<b>HI0855</b> 44%:56%	<i>CHAC1</i> (F) [58]	15:41247844	Frame- shift	p.-177?	-/T	None	23, 19	0.941	0.009	0.997	0	0.787/0
	<i>MIDN</i> (F) [59]	19:1250353	Non-synon	p.C19G	Tgc/Ggc	Missense	3, 7	0.793	0.008	0.998	0	17.85/-3
	<i>CDH6</i> (P) [60]	5:31323179	Non-synon	p.R712G	Aga/Gga	Missense	76, 81	0.749	0.003	0.988	0.01	0.787/-2
<b>HI0868</b> 58%:42%	<i>CHST3</i> (P) [61]	10:73767391	Non-synon	p.Y200C	tAc/tGc	Missense	43, 30	0.759	0	0.998	0.02	17.489/-2
	<i>FBXO5</i> (P) [62]	6:153296090	Non-synon	p.H256R	cAt/cGt	Missense	103, 76	0.756	0.003	1	0	0.787/0
<b>HI0890/NI</b>	<i>BTA1F1</i> (A) [63]	10:93719892	Non-synon	p.Y414F	tAt/tTt	Missense	29, 16	0.765	0.005	0.997	0	3/3
<b>HI1143/14%:86%</b>	<i>CLPTM1</i> (F) [64]	19:45476442	Non-synon	p.N94I	aAc/aTc	Missense	30, 27	0.744	0.007	1	0	0.787/-3
<b>HI1157</b> 61%:39%	<i>AMIGO1</i> (F) [65]	1:110050220	Non-synon	p.R438W	Cgg/Tgg	Missense	42, 28	0.806	0	0.994	0	0.787/-3
	<i>PEX5</i> (P) [66]	12:7362354	Non-synon	p.R560C	Cgc/Tgc	Missense	71, 50	0.801	0.006	1	0	0.787/-3
	<i>KIF23</i> (P) [67]	15:69718474	Non-synon	p.E266A	gAa/gCa	Missense	52, 48	0.775	0.007	0.998	0	0.787/-1
<b>HI1228</b> 42%:58%	<i>FOXA3</i> (P) [68]	19:46376266	Non-synon	p.G334R	Gga/Aga	Missense	30, 29	0.801	0.004	0.975	0	26.000/-2
	<i>FRMD4A</i> (F) [69]	10:13698719	Non-synon	p.S941L	tCg/tTg	Missense	6, 8	0.775	0	0.957	0	0.064/-2

Table 1. Cont.

ID Number	Gene Symbol (Category) *	Chromosome Position (Hg19)	Genomic Variant					Classification Score	GERP2	PolyPhen2 Score	SIFT Score	Blosum Score/Blosum62
			SNP Effect	Amino Acid	SNP Codon	SNP Function	VCF Allele Depth					
<b>HI1305</b> 73%:27%	<i>PIR</i> (F) [70]	X:15474053	Non-synon	p.E132G	gAa/gGa	Missense	32, 28	0.811	0.007	0.962	0	0.213/−2
	<i>DNAH2</i> (F) [71]	17:7637815	Non-synon	p.I255T	aTa/aCa	Missense	25, 20	0.793	0.010	0.942	0	7.006/−1
<b>HI1375</b> 17%:83%	<i>LHX2</i> (F) [72]	9:126777468	Non-synon	p.A130P	Gct/Cct	Missense	50, 34	0.775	0.006	0.998	0	0.787/−1
	<i>EPHA4</i> (P) [73]	2:222428829	Non-synon	p.A148T	Gct/Act	Missense	18, 25	0.744	0	0.977	0	0.787/0
<b>HI1402/85%:15%</b>	<i>ILIRAPL1</i> (A) [12,74]	X:29973282	Non-synon	p.P478Q	cCa/cAa	Missense	57, 42	0.761	0	0.999	0	7.684/−1
<b>HI1422</b> 56%:44%	<i>RTN4R</i> (F) [75]	22:20230387	Non-synon	p.S89L	tCg/tTg	Missense	11, 18	0.778	0.004	0.998	0	0.064/−4
	<i>DHX8</i> (F) [76]	17:41606990	Non-synon	p.R282H	cGc/cAc	Missense	4, 8	0.724	0.010	1	0	0.787/0
<b>HI1433</b> 93%:7%	<i>FAT2</i> (P) [77–79]	5:150947258	Non-synon	p.T411I	aCt/aTt	Missense	32, 29	0.757	0	0.992	0	6.276/−1
	<i>FITM2</i> (F) [80,81]	20:42939631	Non-synon	p.R52H	cGc/cAc	Missense	6, 5	0.750	0.008	0.999	0.02	8.723/0
	<i>RALGPS2</i> (F) [82,83]	1:179064186	Non-synon	p.V342M	Gtg/Atg	Missense	69, 73	0.734	0.010	0.970	0	0.787/1
<b>HI1739</b> 31%:69%	<i>ASB3</i> (F) [84]	2:53941656	Non-synon	p.S208N	aGc/aAc	Missense	53, 30	0.773	0.007	0.980	0	0.403/1
	<i>AKAP6</i> (F) [85]	14:33291533	Non-synon	p.D1504P	gAt/gTt	Missense	43, 41	0.761	0	0.983	0.01	0.787/−3
	<i>GRM4</i> (A) [86]	6:34100940	Non-synon	p.D111N	Gac/Aac	Missense	58, 51	0.726	0.005	0.993	0.02	0.787/1
<b>HI1884/48%:52%</b>	<i>TRAF3IP1</i> (F) [87]	2:239306221	Non-synon	p.M537T	aTg/aCg	Missense	13, 19	0.759	0.007	0.999	0.01	0.033/−1
<b>HI1954</b> 23%:77%	<i>DDX5</i> (P) [88]	17:62500102	Non-synon	p.S146C	tCt/tGt	Missense	88, 91	0.788	0.007	0.950	0	0.787/−1
	<i>KIF5A</i> (P) [89]	12:57974875	Non-synon	p.R891Q	cGg/cAg	Missense	25, 19	0.767	0.005	0.989	0	0.787/1
<b>HI2126/41%:59%</b>	<i>ANKRD11</i> (A) [90,91]	16:89351565	Non-synon	p.T461R	aCa/aGa	Missense	20, 11	0.767	0	0.961	0	0.787/−1
<b>HI2172</b> 61%:39%	<i>TNFRSF21</i> (F) [92]	6:47221105	Non-synon	p.W465R	Tgg/Agg	Missense	8, 16	0.805	0.004	1	0	10.610/−3
	<i>LRSAM1</i> (F) [93,94]	9:130253524	Non-synon	p.L484M	Ctg/Atg	Missense	18, 23	0.796	0.010	0.935	0	0.787/−1
	<i>CYFIP1</i> (A) [95]	15:22960872	Non-synon	p.D716N	Gac/Aac	Missense	28, 30	0.767	0.009	0.999	0	0.067/1
<b>HI2215</b> 60%:40%	<i>FPGT- TNNI3K</i> (F) [96]	1:74836022	Non-synon	p.G673V	gGc/gTc	Missense	34, 53	0.802	0.009	0.998	0	0.787/−3
	<i>MAP1A</i> (P) [96]	15:43820377	Non-synon	p.P2235S	Ccc/Tcc	Missense	65, 67	0.793	0	0.998	0	0.787/−1
	<i>MAP1A</i> (P) [96]	15:43819396	Non-synon	p.Y1908H	Tac/Cac	Missense	13, 21	0.782	0.004	0.994	0	0.787/2

Table 1. Cont.

ID Number	Gene Symbol (Category) *	Chromosome Position (Hg19)	Genomic Variant					Classification Score	GERP2	PolyPhen2 Score	SIFT Score	Blosum Score/Blosum62
			SNP Effect	Amino Acid	SNP Codon	SNP Function	VCF Allele Depth					
<b>HI2244</b> 32%:68%	<i>KDM5B</i> (P) [97–99]	1:202724501	Non-synon	p.Y320K	aTa/aAa	Missense	52, 47	0.808	0.007	0.988	0	0.787/–4
	<i>TRPM2</i> (P) [97–99]	21: 45826549	Non-synon	p.I954F	Atc/Ttc	Missense	17, 16	0.786	0.008	0.997	0	12.615/–3
	<i>RAI1</i> (A) [100]	17:17699284	Non-synon	p.P959A	Ccc/Gcc	Missense	37, 26	0.763	0	0.992	0	0.787/–1
<b>HI2278</b> 8%:92%	<i>DCHS1</i> (F) [39]	11:6662745	Codon Change Plus Codon Insertion	p.L32LW	Ctg/ctCTGg	None	12, 11	0.700	0	0.979	0.02	0/0
<b>HI2843</b> 25%:75%	<i>UBE2E2</i> (P) [101]	3:23631311	Non-synon	p.Y198H	Tac/Cac	Missense	64, 33	0.787	0	0.994	0	0.787/2
	<i>PCDHB15</i> (P) [45]	5:140625618	Non-synon	p.R157W	Cgg/Tgg	Missense	33, 61	0.754	0.007	0.942	0	24.449/–3
	<i>SCG2</i> (F) [102]	2:224463227	Non-synon	p.E257D	gaG/gaC	Missense	70, 64	0.750	0	0.993	0.01	0.787/2
<b>HI2879</b> 18%:82%	<i>TRIM9</i> (F) [103]	14:51448554	Non-synon	p.R623Q	cGg/cAg	Missense	64, 69	0.774	0	1	0	0.787/1
	<i>SOX7</i> (A) [104]	8:10583274	Non-synon	p.Y432H	Tac/Cac	Missense	25, 33	0.766	0	0.997	0	0.787/2
	<i>ANKFN1</i> (P) [105]	17:54559849	Non-synon	p.G744R	Ggg/Agg	Missense	28, 16	0.715	0	0.995	0.04	26.000/–2
	<i>DCHS1</i> (F) [39]	11:6662745	Codon Change Plus Codon Insertion	p.L32LW	ctg/ctCTGg	None	11, 12	0.700	0	0.979	0.02	0/0
<b>HI2898</b> 85%:15%	<i>MPHOSPH8</i> (F) [106]	13:20242552	Non-synon	p.Y736C	tAc/tGc	Missense	39, 28	0.779	0.007	0.999	0	0.787/–2
	<i>GABRQ</i> (A) [107]	X:151819020	Non-synon	p.S292F	tCc/tTc	Missense	9, 16	0.749	0.008	0.963	0	0.787/–2

ID Number represents the AGRE identifier. For the 30 females with ASD in this table, the average age ± standard deviation was 7.7 ± 2.6 years and age range was 5–16 years. Final selection criteria: Classification Score > 0.7; GERP2 < 0.01; PolyPhen2 > 0.9; SIFT < 0.03. \* (A) = Known clinically relevant gene for autism; (F) = Neurodevelopment functional gene for autism; (P) = Known paralogue of an autism gene; NI = not informative; XCI = X chromosome Inactivation (%); ? indicates an unspecified amino acid variation; [ ] represents literature citations for each gene description.

**Table 2.** Summary of Putative Disease Causing Gene Variants Identified by Exome Sequencing.

Gene Variant Category	Number of Subjects (%)	Number of Genes (%)	Range
Known ASD Genes	11 (37%)	12 (15%)	0–2
Paralogues of Known ASD Genes	16 (53%)	23 (30%)	0–2
Neurodevelopmental Function Genes	23 (77%)	43 (54%)	0–7

Six females (HI0605, HI2898, HI1739, HI2172, HI2244 and HI2879) possessed a single variant of a known candidate ASD gene meeting our final selection criteria for pathogenicity in addition to one or more paralogues of ASD genes or gene variants with functional influence over cell growth and neurodevelopment. Fourteen females (HI0714, HI0751, HI0765, HI0855, HI0868, HI1157, HI1228, HI1305, HI1375, HI1422, HI1433, HI1954, HI2215 and HI2843) possessed 2 to 4 variants of ASD gene paralogues or variants with functional roles of influence over ASD but without a conclusive or definable causal mutation. Variants of multiple putative candidate genes for ASD were identified for Subjects HI0405, HI0558 and HI0793.

Using XCI data, we found that eight of the 30 females had high skewness ( $XCI > 80\%:20\%$ ) and an additional six females showed moderate skewness ( $XCI 65\%:35\%$  to  $80\%:20\%$ ). Two of the eight females with high skewness and two females with moderate skewness possessed an X-linked gene variant meeting the final selection criteria. When considering the 78 total genes identified meeting our final selection criteria, X-linked gene variants were disproportionately more likely to be found among the females with autism spectrum disorder who exhibited moderate to high level XCI skewness than females without XCI skewness ( $OR = 6.4$ ,  $95\% CI: 0.68, 60.5$ ,  $p = 0.1$ ) but this relationship did not achieve statistical significance. Similarly, females who exhibited XCI skewness were more likely to possess an X-linked gene variant than females without XCI skewness, ( $OR = 6.0$ ,  $95\% CI: 0.58, 61.8$ ,  $p = 0.132$ ).

### 3. Discussion

We chose females with ASD to study from the age of 5 to 16 years from multiplex families with more than one affected family member in order to increase the likelihood of identifying a single gene as a causative factor using a descriptive case study approach. The females were selected from the AGRE, a biorepository of specimens and clinical data from well-characterized families with children with ASD enrolled for genetic research studies. Females with ASD from multiplex families were also examined to determine whether an overabundance of X-linked genes could be identified contributing to ASD by using whole exome sequencing and XCI assays and if X-linked genes were more frequently disturbed in those females with XCI skewness.

#### 3.1. Single Gene Variants

Our investigation identified four females (HI0555, HI0890, HI1402 and HI2126) with single gene variants of clinically relevant candidate or known genes for ASD with a high likelihood of causality. Subject HI0555 possessed a non-synonymous, missense mutation of the *SETD2* gene located on chromosome 3 with high likelihood to have a deleterious effect on gene expression and/or function.

The *SETD2* encodes a Huntingtin-interacting protein B related to Huntington disease and known expression in the brain [27,28]. The *SETD2* gene is known as an autism susceptibility gene reported in autism [12] and was the highest rated and most likely candidate gene in this case. Subject HI0890 possessed a non-synonymous, missense mutation of the *BTAF1* gene located on chromosome 10 which is implicated as a helicase and sequence-specific binding transcription factor activity [63]. Subject HI2126 possessed a non-synonymous, missense mutation in a known candidate gene for ASD, *ANKRD11* which is located on chromosome 16 and likely to be causative in this case. The *ANKRD11* gene encodes an ankryin repeat domain-containing protein which inhibits ligand-dependent activation of transcription with mutations causing the KBG syndrome characterized by craniofacial features, short stature, skeletal anomalies, seizures and intellectual disability [90,91]. No physical examination or cognitive data were available from AGRE on this 12 year old female diagnosed with autism.

### 3.2. Involvement of Skewness of X Chromosome Inactivation and Putative Disease Causing Genes

High XCI skewness (>80%:20%) was observed for eight of our females with autism and two of these females possessed putative variants of X-linked genes (HI1402 and HI2898) which may contribute to the phenotype. Skewing of X-chromosome inactivation may increase expression of recessive disease causing variants on the X-chromosome. Subject HI1402 with an XCI status of 85%:15% indicating high skewness also possessed a non-synonymous, missense mutation of an X-linked gene (*ILIRAPLI*) which is a recognized candidate gene for ASD [12]. *ILIRAPLI* had the highest criteria selection score of the four genes meeting initial selection criteria and the most likely cause of ASD in this female. The *ILIRAPLI* gene encodes a protein with the highest expression in brain neurons and participates in the regulation of neurite outgrowth via interaction with neuronal calcium sensors thereby regulating synaptic formation and modulation of synaptic transmission [74].

Subject HI2898 with an XCI status of 85%:15% showing high skewness possessed a variant of the X-linked gene (*GABRQ*) found to be in the top two genes in the selection process. The *GABRQ* (gamma-aminobutyric acid A receptor theta) gene encodes a receptor protein for GABAergic neurotransmission in the mammalian central nervous system [107]. GABA is a major inhibitory neurotransmitter. *GABRQ* expression is distributed in the amygdala, hippocampus, anterior hypothalamus and cortex [107]. Subject HI2278 with an XCI status of 8%:92% possessed an X-linked gene (*PAK3*) in the top 25 genes but was excluded due to strict filtering requirements (*i.e.*, PolyPhen 2) and a correct classification score of >0.50. However, the *DCHS1* gene was identified which encodes a transmembrane cell adhesion molecule that belongs to the protocadherin superfamily and acts as a ligand for FAT4, another protocadherin protein. Both *DCHS1* and *FAT4* form an apically located adhesive complex in the developing brain [39].

The five remaining females (HI0751, HI1143, HI1375, HI1433 and HI2879) with high XCI skewness ranging from 93%:7% to 82%:18% showed no X-linked gene variants meeting the final selection criteria for pathogenicity. Other variations of possible clinical relevance in autosomal chromosomes were observed for these subjects. Subject HI1433 with an XCI status of 93%:7% indicating extreme skewness possessed a variant of the *FAT2* gene located on chromosome 5 which is a member of the cadherin-related FAT tumor suppressor homolog 2 (*Drosophila*) [77–79]. Subject HI2879 with an XCI status of 18%:82% showing high skewness possessed a variant of the *SOX7*

(SRY-Box 7) gene which encodes a SOX protein, acting as a transcription factor regulating diverse developmental processes [104]. Based upon our investigation, XCI skewness in these females did not contribute to the presentation of autism.

Moderate skewness (65%:35% to 80%:20%) was observed for six females (HI0765, HI1305, HI1739, HI1954, HI2244 and HI2843). Two of these females possessed putative variants of X-linked genes that may have contributed to their phenotype. Subject HI1305 with an XCI status of 73%:27% possessed a variant for *PIR*, a highly conserved X-linked iron-binding nuclear protein gene which functions as a transcriptional co-regulator and contributes to regulation of cellular processes and may promote apoptosis when over expressed [70]. An associated disorder for this gene when disturbed is extratemporal epilepsy. Three genes were identified for Subject HI0765 with an XCI status of 77%:23% with moderate skewness including *GPRASP2* located on the X chromosome which encodes a protein that may regulate a variety of G-protein coupled receptors associated with autism spectrum disorders and schizophrenia [48], as an X-linked gene approaching high skewness becomes an important candidate for causation of ASD in this female. The *KCNC2* and *ASPM* genes were the highest rated and became likely candidates

### 3.3. Other Autosomal Putative Disease Causing Genes

The cadherin, protocadherin and ankyrin repeat gene families were the most commonly altered putative disease causing gene variants identified in our study. Subject HI0558 possessed a variant of *PCDH8*, a member of the cadherin superfamily of genes and functions in cell adhesion in a CNS-specific manner possibly playing a role in activity-induced synaptic reorganization underlying memory with down-regulation of dendritic spines, primarily in the hippocampal area [35]. In addition this affected female possessed a variant of the *CTNNA3* gene which is located on chromosome 10 and encodes an alpha-t-catenin which plays a role in functional cadherin-mediated cell adhesion. A possible association of this gene in Alzheimer disease has been proposed [36]. Subject HI0855 possessed a variant of *CDH6* located on chromosome 5 encoding a known cadherin playing a role in cell-cell adhesion and implicated in autism [60]. Subject HI0605 possessed a non-synonymous, missense mutation of the *CCDC64* (coiled-coil domain containing, 64), a recognized ASD gene on chromosome 12 [12,38]. *CCDC64* is a component of the secretory vesicle machinery in developing neurons that acts as a regulator of neurite outgrowth in the early phase of neuronal differentiation which when disturbed is associated with neuronitis [38]. Subject HI1739 possessed a non-synonymous missense mutation of the *ASB3* gene located on chromosome 2 and related to the ankyrin repeat gene family known to play a role in brain development and function including autism [88]. *GRM4* (glutamate receptor, metabotropic, 4) is a known ASD gene involved with glutamate, a major excitatory neurotransmitter in the central nervous system [86]. Subject HI2172 possessed a non-synonymous, missense mutation of the *CYFIP1* (cytoplasmic FMRP-interacting protein 1) gene located on chromosome 15 which encodes a protein that interacts with the familial mental retardation protein (FMRP) that when disturbed, causes the fragile X syndrome by impacting on development and maintenance of neuronal structures [108,109]. Individuals with the 15q11.2 BP1-BP2 microdeletion or Burnside-Butler syndrome are known to have developmental and speech delay involving the *CYFIP1* gene with an increased rate of aberrant behavior and autism [95].

## 4. Experimental Section

### 4.1. Samples from Females with Autism

Thirty Caucasian females (average age,  $7.7 \pm 2.6$  years; age range, 5 to 16 years) were selected with confirmed diagnosis of autism. We chose to undertake whole exome sequencing from well-characterized females with autism with a positive family history (e.g., affected brothers) classified as multiplex families with autism and having a high probability of causation due to gene disturbances. We selected affected females from the Autism Genetic Research Exchange (AGRE) repository ([www.agre.autismspeaks.org](http://www.agre.autismspeaks.org)) for a descriptive next generation whole exome sequencing case study. The family members with autism in the AGRE were recruited and enrolled after screening and diagnostic assessments were performed with autism-related testing instruments (e.g., ADOS). Medical examinations and neurology evaluations were undertaken to collect family, pregnancy and medical history data along with physical and anthropometric measures and neurological recordings. Blood was also collected at AGRE for lymphoblastoid cell line development, plasma storage and DNA isolation. Normal chromosome analysis (karyotype) and fragile X syndrome DNA testing were completed previously in all female subjects selected from AGRE. DNA from parents or other family members were not analyzed as a component of our study.

### 4.2. Whole Exome Sequencing

#### 4.2.1. Exome Sequencing Methods

Genomic DNA (5  $\mu$ g) samples were sent to the Silicon Valley Biosystems, Foster City, CA, USA for whole exome sequencing via paired-end next generation sequencing (NGS) approach using standard protocols with the Illumina HiSeq2000 platform (<http://www.service.Illumina.com>) and Agilent SureSelect Human All Exon v4–51Mb (<http://www.genomics.agilent.com>). The primary sequence raw data produced by the sequencing phase of our study was aligned to reference, the variants called and functional significance of each variant determined and rank-ordered into a list of functional variants (in order of decreasing pathology) that were correlated with or causative for the autism phenotype. The raw data was in the form of *fastq* files, containing the sequence reads and quality scores of the NGS sequencing runs. The individual reads were initially aligned by mapping to the reference genome using Burrows-Wheeler aligner (BWA), a software program for mapping low-divergent genomic sequences of up to 100 base pairs against a large reference genome (<http://www.bio-bwa.sourceforge.net>). Variant calling procedures then identified genomic sequence variants from the NGS sequencing runs or the absence of variants in the sample of interest relative to a reference sequence. The platform used to produce variant calls was an amalgamation of tools and methods uniquely combined and specifically calibrated for enhanced performance incorporating *Genome Analysis Toolkit (GATK) V2.39* (<https://www.broadinstitute.org/gatk/>). A number of intermediate steps were applied to improve the quality/accuracy of the alignment and variant calling results prior to the creation of the final list of variants. Sequence reads that were likely to represent duplicates were marked to be ignored by the downstream variant-calling tools. Additionally, realignment around known insertions and deletions was performed, and quality scores recalibrated

based upon the number of reads. The list of variants was produced in the form of a .vcf file annotated to reflect the results of quality control filters. This established and calibrated alignment and variant calling platform performs well and can reliably predict calls for single insertion and deletion events (up to 50 bp with between 97% and 99.9% sensitivity and specificity). The average number of reads per generated fragment was 64 for each of the 30 females with autism.

#### 4.2.2. Data Analysis

The resulting .vcf file contained annotated variants with scientific features, including sequence, gene-, variant-, and transcript-level annotations. Annotations include characteristics of the observed variant such as the genomic coordinates or the variant's genomic region (e.g., exonic, upstream *etc.*), predicted effects on the various transcripts (e.g., missense, frame-shift, *etc.*), observed frequencies of the variant in control populations and affected individuals, curated knowledge from both proprietary and public databases, and pathogenicity predictions (e.g., Polyphen and SIFT). The machine learning tool utilized both the 2012 April 1K Genome release reflecting the frequency at the position and variant levels and the European Exome Sequencing Project (ESP6500) (<https://esp.gs.washington.edu/drupal/>) at the variant level which reflects the frequency of variants as observed in the ESP6500 dataset of healthy individuals in generating the correct classification score. The regulatory significance of variants were derived from ENCODE data. A classification engine developed by Silicon Valley Biosystems was used to create a pathogenicity score based on a machine learning technique approach. The model is trained and validated using 150,000 gene polymorphisms and mutations from a large set of variants with known pathogenicity. The classification accuracy exceeds 99% and thus the score provides the probability that a particular variant has an effect on gene expression or protein structure. The classifier was trained using known disease-causing mutations and known benign polymorphisms, each of which was annotated with attributes such as location relative to splice junctions, type of non-synonymous amino acid substitution and allele frequency in the population. The classifier was validated on a similarly sized set of variants and was shown to perform with 93% sensitive and 97% specificity. Prioritization of putative functionally relevant variants based on the classification score was performed and these variants were additionally classified with other means of determining pathogenicity such as SIFT, PolyPhen2, Blosum and Blosum62. The variants that were the most highly predicted to be functionally significant were examined across the study cohort for clustering within and around genes hypothesized as relevant to the phenotype.

#### 4.2.3. Gene Filtration/Selection Parameters

A high filter system was utilized whereby the indels and quality disqualified (QD) parameters were identified and removed based on the bioinformatics outlined in our study approach. The classification score (range 0 to 1) was generated as our primary delimiting factor in identifying novel or candidate genes for autism in the affected females. The classifier uses machine learning algorithms to predict a variant's functional impact on a gene with higher scores for genomic variants considered more pathogenic or disease-causing changes at the gene level. Thus, we applied an initial classification cutoff of >0.5 with final selection criteria >0.7. Genomic Evolutionary Rate Profiling (GERP2), an indicator of the prevalence of the genomic variants in the general population, was assigned a cutoff

value of  $<0.01$  and remained at  $<0.01$  as final selection criteria. GERP2 identifies constrained elements in multiple alignments by quantifying substitution deficits. These deficits represented substitutions that would have occurred if the element was neutral DNA, but did not occur because the element was under functional constraint. The initial Polymorphism Phenotyping v2 (PolyPhen2) cutoff value was  $>0.85$  with the final selection criteria  $>0.9$  and the initial Sorts Intolerant From Tolerant (SIFT) cutoff value was  $<0.05$  with a final selection criteria  $<0.03$  were predictors of pathogenicity or disease causing changes. The Blosum (BLOcks SUBstitution Matrix) score reports the ratio between the frequency of an amino acid substitution in the Human Gene Mutation Database (HGMD) to the frequency of this amino acid substitution in known polymorphism data sets. Numbers greater than 1 indicate this variant was observed more in diseased individuals than in controls and less likely to be tolerated. We also used the Blosum62 matrix to score alignments between evolutionarily divergent protein sequences based on local alignments. Blosum62 is a log-odds score for each possible substitution of the 20 standard amino acids. Higher Blosum62 numbers represent more closely related species comparisons and provide a general indication of frequency of the substitution. Negative numbers indicate less frequent substitutions and a lower likelihood to be tolerated (e.g., negative values = unexpected or rare substitutions).

The list of disease causing genes were developed using the primary selection criteria (*i.e.*, classification score) and then adding cutoff levels for other predictive parameters (GERP2, PolyPhen2, and SIFT). The initial list of genes and genomic variants used the classification score  $>0.5$  alone and then the cutoff levels of the other predictive parameters (GERP2, PolyPhen2 and SIFT) were included. The classification score was then raised to  $>0.7$  to generate the final list of genes. These putative candidate genes and genomic variants were then subjected to further screening for biological significance for neural development, function and known neurological disorders using Online Mendelian Inheritance in Man (OMIM, <http://www.ncbi.nlm.nih.gov/omim>), GeneCards (<http://www.genecards.org/>), PubMed and other online websites and databases.

#### 4.2.4. Clinical Relevance to ASD

Following the recommendations of the American College of Medical Genetics and Genomics (ACMGG) in reporting clinical exome and genome sequencing data, we analyzed only the primary findings, in depth, which is termed as the pathogenic alteration in a gene or genes relevant to the diagnostic indication for sequencing study, *i.e.*, causation of autism in females. Only clinically relevant candidate or known existing genes for autism were included for extended analysis. Incidental findings are those pathogenic or likely pathogenic alterations in genes that are disease associated or causing but not apparently relevant to a diagnostic indication for which sequencing was undertaken (*i.e.*, in our study on autism [12]) and were not further investigated in our study. We did not include mutations of recognized autosomal recessive genes causing human disorder (e.g., *GJB2* or *GJB6*, gene mutations causing hearing loss). It is estimated that about 1% of sequencing reports would include an incidental variant.

### 4.3. X Chromosome Inactivation in Females with Autism

Genomic DNA isolated from blood was used as a template for polymerase chain reaction (PCR) amplification, to identify the CAG polymorphic region of the *AR* gene. Prior to the PCR amplification, 200 ng of genomic DNA was digested with the methyl-sensitive restriction enzyme *HpaII* [21]. Approximately 50 ng of digested or undigested genomic DNA was used as a template for PCR amplification to determine the peak height of the polymorphic PCR fragment of the *AR* gene using the following primers: forward 5' TCCAGAATCTGTTCCAGAGCGTGC 3' and reverse 5' GCTGTGAA GGTGCTGTTCTCAT 3' with the forward primer fluorescently labeled with 6-FAM. The lengths and peak heights of the resulting PCR fragments are determined with the use of capillary electrophoresis and fragment separation software with the ABI 3100 DNA sequencer (Applied Biosystems, Carlsbad, CA, USA) using established protocols [18,22].

The digestion process preferentially degrades activated (unmethylated) over inactivated (methylated) DNA. Undigested DNA is preferentially amplified and produces larger peak heights. The peak height values for digested DNA is normalized using peak height values for the undigested DNA for each subject. The percentage of XCI for each *AR* allele was then calculated using the following formula:  $(d1/u1)/[(d1/u1) + (d2/u2)]$ ;  $d1$  = peak height of digested DNA from the first allele and  $u1$  = peak height of undigested DNA from the first allele;  $d2$  = peak height of digested DNA from the second allele and  $u2$  = peak height of undigested DNA from the second allele. Highly skewed XCI is defined as >80% calculated ratio for either one of the *AR* gene alleles in the digested DNA sample. To ensure the reproducibility of XCI results and equal amplification of both alleles, the digestion, PCR amplification, and genotyping were repeated up to three times in several samples.

XCI status for each subject was assigned to one of the three mutually exclusive categories: (i) randomly selected inactivation of either allele from each X chromosome (XCI = 50%:50% to 64%:36%); (ii) moderately skewed inactivation favoring 1 allele of one of the X chromosomes (XCI = 65%:35% to 80%:20%); and (iii) highly skewed inactivation of a single allele representing one X chromosome (XCI > 80%:20%). The relative frequency of random, moderate, and highly skewed XCI categories was determined for the females with autism. The binomial frequency distribution of X-linked gene variants (present or absent) among females with autism spectrum disorder exhibiting moderate and high levels of XCI skewness were approximated using Yates' chi-square test and odds ratios were calculated with 95% confidence intervals.

The top genes chosen were based on the final selection criteria identified for each female and evaluated for biologic function and whether each gene was previously reported as a clinically relevant candidate or known gene for autism [12]. The X chromosome inactivation data and XCI status on each female were used to support whether an X-linked gene played a role in the causation of autism in the affected female if XCI skewness was found.

## 5. Conclusions

In summary, we shared our experience using a descriptive next generation whole exome sequencing case study approach examining 30 well-characterized Caucasian females with autism between 5 and 16 years of age recruited from multiplex families enrolled at the AGRE for research purposes.

Interpretation of DNA findings are limited due to an inability to study other affected and non-affected family members to determine whether the gene variants were de-novo or inherited in origin for correlation with clinical phenotypes and lack of Sanger sequencing confirmation. Using strict selection criteria, four females (13%) were found to possess single gene variants of known candidate genes for ASD with a high likelihood of causality. We also identified multiple plausible candidate genes [some known (e.g., *CCDC64* in subject HI0605) and some novel (e.g., *CHAC1* in subject HI0855)] in 77% of the remaining females. In most females, we found more than one gene that could contribute to the causation of autism. This compares with a recent report of molecular findings among 2000 consecutive patients (primarily pediatric age) referred for clinical whole exome sequencing for evaluation of suspected genetic disorders at a large USA medical center in which 25% received a molecular diagnosis. About 60% of the diagnostic mutations were not previously reported [110]. By phenotypic category, 36% of those presenting with neurological involvement were found to have a molecular diagnosis while 20% of those from the non-neurological group had a diagnosis. We not only performed whole exome sequencing in our females with autism but also performed X chromosome inactivation (XCI) studies for evidence of X-linked genes in the females showing non-random XCI skewness. Moderate and high XCI skewness was seen in 14 of the 30 females and those with skewness were 4 times more likely to show a gene variant on the X chromosome.

### Acknowledgments

The authors thank Carla Meister for manuscript preparation. Partial grant support was provided by the University of Kansas Medical Center and Center of Translational Science through the NIH National Center for Advancing Translational Sciences (Grant# UL1TR000001) and from NICHD (HD02528).

### Author Contributions

Merlin G. Butler conceived the study, analyzed data and wrote the manuscript. Syed K. Rafi reviewed literature, assessed the function of gene variants and contributed to the content of the manuscript. Waheeda Hossain carried out DNA isolation and X-chromosome inactivation studies with data analysis. Dietrich A. Stephan carried out whole exome sequencing experiments, data analysis and interpretation and contributed to the content of the manuscript. Ann M. Manzardo carried out and summarized data analysis and interpretation and contributed to the text of the manuscript.

### Conflicts of Interest

The authors declare no conflict of interest.

### References

1. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*, 4th ed.; American Psychiatric Association Press: Washington, DC, USA, 2000.
2. Johnson, C.P.; Myers, S.M.; The Council on Children With Disabilities. Identification and evaluation of children with autism spectrum disorders. *Pediatrics* **2007**, *120*, 1183–1215.

3. Lord, C.; Risi, S.; Lambrecht, L.; Cook, E.H., Jr.; Leventhal, B.L.; DiLavore, P.C.; Pickles, A.; Rutter, M. The autism diagnostic observation schedule-generic: A standard measure of social and communication deficits associated with the spectrum of autism. *J. Autism Dev. Disord.* **2000**, *30*, 205–223.
4. Le Couteur, A., Lord, C., Rutter, M. *Autism Diagnostic Interview-Reviewed (ADI-R)*; Western Psychological Services: Los Angeles, CA, USA, 2003.
5. Constantino, J.N.; Davis, S.A.; Todd, R.D.; Schindler, M.K.; Gross, M.M.; Brophy, S.L.; Metzger, L.M.; Shoushtari, C.S.; Splinter, R.; Reich, W. Validation of a brief quantitative measure of autistic traits: Comparison of the social responsiveness scale with the autism diagnostic interview-revised. *J. Autism Dev. Disord.* **2003**, *33*, 427–433.
6. Developmental Disabilities Monitoring Network Surveillance Year 2010 Principal Investigators; Centers for Disease Control and Prevention (CDC). Prevalence of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, United States, 2010. *MMWR Surveill. Summ.* **2014**, *63*, 1–21.
7. Rapin, I.; Dunn, M. The neurology of autism: Many unanswered questions. *Eur. J. Neurol.* **1995**, *2*, 151–162.
8. Geschwind, D.H.; Levitt, P. Autism spectrum disorders: Developmental disconnection syndromes. *Curr. Opin. Neurobio.* **2007**, *17*, 103–111.
9. Butler, M.G.; Dasouki, M.J.; Zhou, X.P.; Talebizadeh, Z.; Brown, M.; Takahashi, T.N.; Miles, J.H.; Wang, C.H.; Stratton, R.; Pilarski, R.; *et al.* Subset of individuals with autism spectrum disorders and extreme macrocephaly associated with germline PTEN tumour suppressor gene mutations. *J. Med. Genet.* **2005**, *42*, 318–321.
10. Herman, G.E.; Henninger, N.; Ratliff-Schaub, K.; Pastore, M.; Fitzgerald, S.; McBride, K.L. Genetic testing in autism: How much is enough? *Genet. Med.* **2007**, *9*, 268–274.
11. Roberts, J.L.; Hovanes, K.; Dasouki, M.; Manzardo, A.M.; Butler, M.G. Chromosomal microarray analysis of consecutive individuals with autism spectrum disorders or learning disability presenting for genetic services. *Gene* **2014**, *535*, 70–78.
12. Butler MG, Rafi, S.K.; Manzardo, AM. Clinically relevant candidate and known genes for autism spectrum disorders (ASD) with representation on high resolution chromosome ideograms. *OA Autism* **2014**, *2*, 5–28.
13. Lyon, M.F. X-chromosome inactivation and human genetic disease. *Acta Paediatr.* **2002**, *91*, 107–112.
14. Plenge, R.M.; Stevenson, R.A.; Lubs, H.A.; Schwartz, C.E.; Willard, H.F. Skewed X-chromosome inactivation is a common feature of X-linked mental retardation disorders. *Am. J. Hum. Genet.* **2002**, *71*, 168–173.
15. Talebizadeh, Z.; Bittel, D.C.; Veatch, O.J.; Kibiryeva, N.; Butler, M.G. Brief report: Non-random X chromosome inactivation in females with autism. *J. Autism Dev. Disord.* **2005**, *35*, 675–681.
16. Buyse, I.M.; Fang, P.; Hoon, K.T.; Amir, R.E.; Zoghbi, H.Y.; Roa, B.B. Diagnostic testing for Rett syndrome by DHPLC and direct sequencing analysis of the MECP2 gene: Identification of several novel mutations and polymorphisms. *Am. J. Hum. Genet.* **2000**, *67*, 1428–1436.
17. Maier, E.M.; Kammerer, S.; Muntau, A.C.; Wichers, M.; Braun, A.; Roscher, A.A. Symptoms in carriers of adrenoleukodystrophy relate to skewed X inactivation. *Ann. Neurol.* **2002**, *52*, 683–688.

18. Butler, M.G.; Theodoro, M.F.; Bittel, D.C.; Kuipers, P.J.; Driscoll, D.J.; Talebizadeh, Z. X-chromosome inactivation patterns in females with Prader-Willi syndrome. *Am. J. Med. Genet. A* **2007**, *143*, 469–475.
19. Butler, M.G.; Sturich, J.; Myers, S.E.; Gold, J.A.; Kimonis, V.; Driscoll, D.J. Is gestation in Prader-Willi syndrome affected by the genetic subtype? *J. Assist. Reprod. Genet.* **2009**, *26*, 461–466.
20. Cassidy, S.B.; Lai, L.W.; Erickson, R.P.; Magnuson, L.; Thomas, E.; Gendron, R.; Herrmann, J. Trisomy 15 with loss of the paternal 15 as a cause of Prader-Willi syndrome due to maternal disomy. *Am. J. Hum. Genet.* **1992**, *51*, 701–708.
21. Allen, R.C.; Zoghbi, H.Y.; Moseley, A.B.; Rosenblatt, H.M.; Belmont, J.W. Methylation of HpaII and HhaI sites near the polymorphic CAG repeat in the human androgen-receptor gene correlates with x chromosome inactivation. *Am. J. Hum. Genet.* **1992**, *51*, 1229–1239.
22. Bittel, D.C.; Theodoro, M.F.; Kibiryeve, N.; Fischer, W.; Talebizadeh, Z.; Butler, M.G. Comparison of X-chromosome inactivation patterns in multiple tissues from human females. *J. Med. Genet.* **2008**, *45*, 309–313.
23. Bond, J.; Roberts, E.; Mochida, G.H.; Hampshire, D.J.; Scott, S.; Askham, J.M.; Springell, K.; Mahadevan, M.; Crow, Y.J.; Markham, A.F.; *et al.* ASPM is a major determinant of cerebral cortical size. *Nat. Genet.* **2002**, *32*, 316–320.
24. Vermeulen, M.; Mulder, K.W.; Denissov, S.; Pim Pijnappel, W.W. M.; van Schaik, F.M.A.; Varier, R.A.; Baltissen, M.P.A.; Stunnenberg, H.G.; Mann, M.; Timmers, H.T.M. Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell* **2007**, *131*, 58–69.
25. Lacy, S.E.; Bonnemann, C.G.; Buzney, E.A.; Kunkel, L.M. Identification of FLRT1, FLRT2, and FLRT3: A novel family of transmembrane leucine-rich repeat proteins. *Genomics* **1999**, *62*, 417–426.
26. Wang, J.; Takeuchi, T.; Yokota, H.; Izumi, T. Novel rabphilin-3-like protein associates with insulin-containing granules in pancreatic beta cells. *J. Biol. Chem.* **1999**, *274*, 28542–28548.
27. Faber, P.W.; Barnes, G.T.; Srinidhi, J.; Chen, J.; Gusella, J.F.; MacDonald, M.E. Huntingtin interacts with a family of WW domain proteins. *Hum. Mol. Genet.* **1998**, *7*, 1463–1474.
28. Nagase, T.; Kikuno, R.; Hattori, A.; Kondo, Y.; Okumura, K.; Ohara, O. Prediction of the coding sequences of unidentified human genes. XIX. The complete sequences of 100 new cDNA clones from brain which code for large proteins *in vitro*. *DNA Res.* **2000**, *7*, 347–355.
29. Claudio, J.O.; Liew, C.C.; Ma, J.; Heng, H.H.Q.; Stewart, A.K.; Hawley, R.G. Cloning and expression analysis of a novel WD repeat gene, WDR3, mapping to 1p12-p13. *Genomics* **1999**, *59*, 85–89.
30. Iwamoto, M.; Asakawa, H.; Hiraoka, Y.; Haraguchi, T. Nucleoporin Nup98: A gatekeeper in the eukaryotic kingdoms. *Genes Cells* **2010**, *15*, 661–669.
31. Tommerup, N.; Vissing, H. Isolation and fine mapping of 16 novel human zinc finger-encoding cDNAs identify putative candidate genes for developmental and malignant disorders. *Genomics* **1995**, *27*, 259–264.
32. Chapelin, C.; Duriez, N.; Magnino, F.; Goossens, M.; Escudier, E.; Amselem, S. Isolation of several human axonemal dynein heavy chain genes: Genomic structure of the catalytic site, phylogenetic analysis and chromosomal assignment. *FEBS Lett.* **1997**, *412*, 325–330.

33. Bond, J.; Roberts, E.; Springell, K.; Lizarraga, S.B.; Scott, S.; Higgins, J.; Hampshire, D.J.; Morrison, E.E.; Leal, G.F.; Silva, E.O.; *et al.* A centrosomal mechanism involving CDK5RAP2 and CENPJ controls brain size. *Nat. Genet.* **2005**, *37*, 353–355.
34. Pertel, T.; Hausmann, S.; Morger, D.; Zuger, S.; Guerra, J.; Lascano, J.; Reinhard, C.; Santoni, F.A.; Uchil, P.D.; Chatel, L.; *et al.* TRIM5 is an innate immune sensor for the retrovirus capsid lattice. *Nature* **2011**, *472*, 361–365.
35. Strehl, S.; Glatt, K.; Liu, Q.M.; Glatt, H.; Lalande, M. Characterization of two novel protocadherins (PCDH8 and PCDH9) localized on human chromosome 13 and mouse chromosome 14. *Genomics* **1998**, *53*, 81–89.
36. Miyashita, A.; Arai, H.; Asada, T.; Imagawa, M.; Matsubara, E.; Shoji, M.; Higuchi, S.; Urakami, K.; Kakita, A.; Takahashi, H.; *et al.* Genetic association of CTNNA3 with late-onset Alzheimer's disease in females. *Hum. Mol. Genet.* **2007**, *16*, 2854–2869.
37. Kao, S.C.; Chen, C.Y.; Wang, S.L.; Yang, J.J.; Hung, W.C.; Chen, Y.C.; Lai, N.S.; Liu, H.T.; Huang, H.; Chen, H.C.; *et al.* Identification of phostensin, a PP1 F-actin cytoskeleton targeting subunit. *Biochem. Biophys. Res. Commun.* **2007**, *356*, 594–598.
38. Leidinger, P.; Backes, C.; Deutscher, S.; Schmitt, K.; Mueller, S.C.; Frese, K.; Haas, J.; Ruprecht, K.; Paul, F.; Stahler, C.; *et al.* A blood based 12-miRNA signature of Alzheimer disease patients. *Genome Biol.* **2013**, *14*, doi:10.1186/gb-2013-14-7-r78.
39. Cappello, S.; Gray, M.J.; Badouel, C.; Lange, S.; Einsiedler, M.; Srour, M.; Chitayat, D.; Hamdan, F.F.; Jenkins, Z.A.; Morgan, T.; *et al.* Mutations in genes encoding the cadherin receptor-ligand pair DCHS1 and FAT4 disrupt cerebral cortical development. *Nat. Genet.* **2013**, *45*, 1300–1308.
40. Grishin, A.; Li, H.; Levitan, E.S.; Zaks-Makhina, E. Identification of gamma-aminobutyric acid receptor-interacting factor 1 (TRAK2) as a trafficking factor for the K<sup>+</sup> channel Kir2.1. *J. Biol. Chem.* **2006**, *281*, 30104–30111.
41. Billaut-Laden, I.; Rat, E.; Allorge, D.; Crunelle-Thibaut, A.; Cauffiez, C.; Chevalier, D.; Lo-Guidice, J.M.; Broly, F. Evidence for a functional genetic polymorphism of the human mercaptopyruvate sulfurtransferase (MPST), a cyanide detoxification enzyme. *Toxicol. Lett.* **2006**, *165*, 101–111.
42. Hayette, S.; Gadoux, M.; Martel, S.; Bertrand, S.; Tigaud, I.; Magaud, J.P.; Rimokh, R. FLRG (follistatin-related gene), a new target of chromosomal rearrangement in malignant blood disorders. *Oncogene* **1998**, *16*, 2949–2954.
43. Bajorek, M.; Morita, E.; Skalicky, J.J.; Morham, S.G.; Babst, M.; Sundquist, W.I. Biochemical analyses of human IST1 and its function in cytokinesis. *Mol. Biol. Cell.* **2009**, *20*, 1360–1373.
44. Klejnot, M.; Kozielski, F. Structural insights into human Kif7, a kinesin involved in hedgehog signalling. *Acta Crystallogr.* **2012**, *68*, 154–159.
45. Wu, Q.; Maniatis, T. A striking organization of a large family of human neural cadherin-like cell adhesion genes. *Cell* **1999**, *97*, 779–790.
46. Tran, T.S.; Rubio, M.E.; Clem, R.L.; Johnson, D.; Case, L.; Tessier-Lavigne, M.; Haganir, R.L.; Ginty, D.D.; Kolodkin, A.L. Secreted semaphorins control spine distribution and morphogenesis in the postnatal CNS. *Nature* **2009**, *462*, 1065–1069.

47. Zhang, X.Y.; Wen, J.; Yang, W.; Wang, C.; Gao, L.; Zheng, L.H.; Wang, T.; Ran, K.; Li, Y.; Li, X.; *et al.* Gain-of-function mutations in SCN11a cause familial episodic pain. *Am. J. Hum. Genet.* **2013**, *93*, 957–966.
48. Piton, A.; Gauthier, J.; Hamdan, F.F.; Lafreniere, R.G.; Yang, Y.; Henrion, E.; Laurent, S.; Noreau, A.; Thibodeau, P.; Karemera, L.; *et al.* Systematic resequencing of X-chromosome synaptic genes in autism spectrum disorder and schizophrenia. *Mol. Psychiatry* **2011**, *16*, 867–880.
49. Agrawal, A.; Pergadia, M.L.; Saccone, S.F.; Lynskey, M.T.; Wang, J.C.; Martin, N.G.; Statham, D.; Henders, A.; Campbell, M.; Garcia, R.; *et al.* An autosomal linkage scan for cannabis use disorders in the nicotine addiction genetics project. *Arch. Gen. Psychiatry* **2008**, *65*, 713–721.
50. Yoshimura, S.; Gerondopoulos, A.; Linfood, A.; Rigden, D.J.; Barr, F.A. Family-wide characterization of the DENN domain Rab GDP-GTP exchange factors. *J. Cell Biol.* **2010**, *191*, 367–381.
51. Le Meur, N.; Martin, C.; Saugier-veber, P.; Joly, G.; Lemoine, F.; Moirrot, H.; Rossi, A.; Bachy, B.; Cabot, A.; Joly, P.; *et al.* Complete germline deletion of the *STK11* gene in a family with Peutz-Jeghers syndrome. *Eur. J. Hum. Genet.* **2004**, *12*, 415–418.
52. Smolen, G.A.; Schott, B.J.; Stewart, R.A.; Diederichs, S.; Muir, B.; Provencher, H.L.; Look, A.T.; Sgroi, D.C.; Peterson, R.T.; Haber, D.A. A Rap GTPase interactor, RADIL, mediates migration of neural crest precursors. *Genes Dev.* **2007**, *21*, 2131–2136.
53. Zuo, X.; Zhang, J.; Zhang, Y.; Hsu, S.C.; Zhou, D.; Guo, W. Exo70 interacts with the Arp2/3 complex and regulates cell migration. *Nat. Cell Biol.* **2006**, *8*, 1383–1388.
54. Oeffner, F.; Moch, C.; Neundorff, A.; Hofmann, J.; Koch, M.; Grzeschik, K.H. Novel interaction partners of Bardet-Biedl syndrome proteins. *Cell Motil. Cytoskelet.* **2008**, *65*, 143–155.
55. Lopez, I.; Mak, E.C.; Ding, J.; Hamm, H.E.; Lomasney, J.W. A novel bifunctional phospholipase C that is regulated by G $\alpha$ 12 and stimulates the Ras/mitogen-activated protein kinase pathway. *J. Biol. Chem.* **2001**, *276*, 2758–2765.
56. Liefke, R.; Oswald, F.; Alvarado, C.; Ferres-Marco, D.; Mittler, G.; Rodriguez, P.; Dominguez, M.; Borggrefe, T. Histone demethylase KDM5A is an integral part of the core Notch-RBP-J repressor complex. *Genes Dev.* **2010**, *24*, 590–601.
57. Chen, P.W.; Jian, X.; Yoon, H.Y.; Randazzo, P.A. ARAP2 signals through Arf6 and Rac1 to control focal adhesion morphology. *J. Biol. Chem.* **2013**, *288*, 5849–5860.
58. Chi, Z.; Zhang, J.; Tokunaga, A.; Harraz, M.M.; Byrne, S.T.; Dolinko, A.; Xu, J.; Blackshaw, S.; Gaiano, N.; Dawson, T.M.; *et al.* Botch promotes neurogenesis by antagonizing Notch. *Dev. Cell* **2012**, *22*, 707–720.
59. Tsukahara, M.; Suemori, H.; Noguchi, S.; Ji, Z.S.; Tsunoo, H. Novel nucleolar protein, midnolin, is expressed in the mesencephalon during mouse development. *Gene* **2000**, *254*, 45–55.
60. Suzuki, S.; Sano, K.; Tanihara, H. Diversity of the cadherin family: Evidence for eight new cadherins in nervous tissue. *Cell Regul.* **1991**, *2*, 261–270.
61. Unger, S.; Lausch, E.; Rossi, A.; Megarbane, A.; Sillence, D.; Alcausin, M.; Aytes, A.; Mendoza-Londono, R.; Nampoothiri, S.; Afroze, B.; *et al.* Phenotypic features of carbohydrate sulfotransferase 3 (CHST3) deficiency in 24 patients: Congenital dislocations and vertebral changes as principal diagnostic features. *Am. J. Med. Genet. A* **2010**, *152A*, 2543–2549.

62. Reimann, J.D.; Freed, E.; Hsu, J.Y.; Kramer, E.R.; Peters, J.M.; Jackson, P.K. Emi1 is a mitotic regulator that interacts with Cdc20 and inhibits the anaphase promoting complex. *Cell* **2001**, *105*, 645–655.
63. Van Der Knaap, J.A.; van Den Boom, V.; Kuipers, J.; van Eijk, M.J.; van Der Vliet, P.C.; Timmers, H.T. The gene for human TATA-binding-protein-associated factor (TAFII) 170: Structure, promoter and chromosomal localization. *Biochem. J.* **2000**, *345*, 521–527.
64. Yoshiura, K.; Machida, J.; Daack-Hirsch, S.; Patil, S.R.; Ashworth, L.K.; Hecht, J.T.; Murray, J.C. Characterization of a novel gene disrupted by a balanced chromosomal translocation t(2;19)(q11.2;q13.3) in a family with cleft lip and palate. *Genomics* **1998**, *54*, 231–240.
65. Kuja-Panula, J.; Kiiltomaki, M.; Yamashiro, T.; Rouhiainen, A.; Rauvala, H. AMIGO, a transmembrane protein implicated in axon tract development, defines a novel protein family with leucine-rich repeats. *J. Cell Biol.* **2003**, *160*, 963–973.
66. Dammai, V.; Subramani, S. The human peroxisomal targeting signal receptor, Pex5p, is translocated into the peroxisomal matrix and recycled to the cytosol. *Cell* **2001**, *105*, 187–196.
67. Iwamori, T.; Iwamori, N.; Ma, L.; Edson, M.A.; Greenbaum, M.P.; Matzuk, M.M. TEX14 interacts with CEP55 to block cell abscission. *Mol. Cell. Biol.* **2010**, *30*, 2280–2292.
68. Friedman, J.R.; Kaestner, K.H. The Foxa family of transcription factors in development and metabolism. *Cell Mol. Life Sci.* **2006**, *63*, 2317–2328.
69. Yoon, D.; Kim, Y.J.; Cui, W.Y.; Van der Vaart, A.; Cho, Y.S.; Lee, J.Y.; Ma, J.Z.; Payne, T.J.; Li, M.D.; Park, T. Large-scale genome-wide association study of asian population reveals genetic factors in FRMD4A and other loci influencing smoking initiation and nicotine dependence. *Hum. Genet.* **2012**, *131*, 1009–1021.
70. Licciulli, S.; Cambiaghi, V.; Scafetta, G.; Gruszka, A.M.; Alcalay, M. Pirin downregulation is a feature of AML and leads to impairment of terminal myeloid differentiation. *Leukemia* **2010**, *24*, 429–437.
71. Pazour, G.J.; Agrin, N.; Walker, B.L.; Witman, G.B. Identification of predicted human outer dynein arm genes: Candidates for primary ciliary dyskinesia genes. *J. Med. Genet.* **2006**, *43*, 62–73.
72. Chou, S.J.; O’Leary, D.D. Role for Lhx2 in corticogenesis through regulation of progenitor differentiation. *Mol. Cell. Neurosci.* **2013**, *56*, 1–9.
73. Leighton, P.A.; Mitchell, K.J.; Goodrich, L.V.; Lu, X.; Pinson, K.; Scherz, P.; Skarnes, W.C.; Tessier-Lavigne, M. Defining brain wiring patterns and mechanisms through gene trapping in mice. *Nature* **2001**, *410*, 174–179.
74. Piton, A.; Michaud, J.L.; Peng, H.; Aradhya, S.; Gauthier, J.; Mottron, L.; Champagne, N.; Lafreniere, R.G.; Hamdan, F.F.; team, S.D.; *et al.* Mutations in the calcium-related gene IL1RAPL1 are associated with autism. *Hum. Mol. Genet.* **2008**, *17*, 3965–3974.
75. Sinibaldi, L.; De Luca, A.; Bellacchio, E.; Conti, E.; Pasini, A.; Paloscia, C.; Spalletta, G.; Caltagirone, C.; Pizzuti, A.; Dallapiccola, B. Mutations of the Nogo-66 receptor (RTN4R) gene in schizophrenia. *Hum. Mutat.* **2004**, *24*, 534–535.
76. Ono, Y.; Ohno, M.; Shimura, Y. Identification of a putative RNA helicase (HRH1), a human homolog of yeast Prp22. *Mol. Cell. Biol.* **1994**, *14*, 7611–7620.

77. Katoh, Y.; Katoh, M. Comparative integromics on FAT1, FAT2, FAT3 and FAT4. *Int. J. Mol. Med.* **2006**, *18*, 523–528.
78. Neale, B.M.; Kou, Y.; Liu, L.; Ma'ayan, A.; Samocha, K.E.; Sabo, A.; Lin, C.F.; Stevens, C.; Wang, L.S.; Makarov, V.; *et al.* Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* **2012**, *485*, 242–245.
79. Kenny, E.M.; Cormican, P.; Furlong, S.; Heron, E.; Kenny, G.; Fahey, C.; Kelleher, E.; Ennis, S.; Tropea, D.; Anney, R.; *et al.* Excess of rare novel loss-of-function variants in synaptic genes in schizophrenia and autism spectrum disorders. *Mol. Psychiatry* **2014**, *19*, 872–879.
80. Kadereit, B.; Kumar, P.; Wang, W.J.; Miranda, D.; Snapp, E.L.; Severina, N.; Torregroza, I.; Evans, T.; Silver, D.L. Evolutionarily conserved gene family important for fat storage. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 94–99.
81. Bai, S.W.; Herrera-Abreu, M.T.; Rohn, J.L.; Racine, V.; Tajadura, V.; Suryavanshi, N.; Bechtel, S.; Wiemann, S.; Baum, B.; Ridley, A.J. Identification and characterization of a set of conserved and new regulators of cytoskeletal organization, cell morphology and migration. *BMC Biol.* **2011**, *9*, 54.
82. Liu, F.; Arias-Vasquez, A.; Sleegers, K.; Aulchenko, Y.S.; Kayser, M.; Sanchez-Juan, P.; Feng, B.J.; Bertoli-Avella, A.M.; van Swieten, J.; Axenovich, T.I.; *et al.* A genomewide screen for late-onset Alzheimer disease in a genetically isolated Dutch population. *Am. J. Hum. Genet.* **2007**, *81*, 17–31.
83. Jin, J.; Smith, F.D.; Stark, C.; Wells, C.D.; Fawcett, J.P.; Kulkarni, S.; Metalnikov, P.; O'Donnell, P.; Taylor, P.; Taylor, L.; *et al.* Proteomic, functional, and domain-based analysis of *in vivo* 14–3–3 binding proteins involved in cytoskeletal regulation and cellular organization. *Curr. Biol.* **2004**, *14*, 1436–1450.
84. Kile, B.T.; Viney, E.M.; Willson, T.A.; Brodnicki, T.C.; Cancilla, M.R.; Herlihy, A.S.; Croker, B.A.; Baca, M.; Nicola, N.A.; Hilton, D.J.; *et al.* Cloning and characterization of the genes encoding the ankyrin repeat and SOCS box-containing proteins asb-1, asb-2, asb-3 and asb-4. *Gene* **2000**, *258*, 31–41.
85. Dodge-Kafka, K.L.; Soughayer, J.; Pare, G.C.; Michel, J.J.C.; Langeberg, L.K.; Kapiloff, M.S.; Scott, J.D. The protein kinase A anchoring protein mAKAP coordinates two integrated cAMP effector pathways. *Nature* **2005**, *437*, 574–578.
86. Wu, S.; Wright, R.A.; Rockey, P.K.; Burgett, S.G.; Arnold, J.S.; Rosteck, P.R., Jr.; Johnson, B.G.; Schoepp, D.D.; Belagaje, R.M. Group III human metabotropic glutamate receptors 4, 7 and 8: Molecular cloning, functional expression, and comparison of pharmacological properties in RGT cells. *Brain Res. Mol. Brain Res.* **1998**, *53*, 88–97.
87. Barbari, N.F.; Kin, N.W.; Sharma, N.; Michaud, E.J.; Kesterson, R.A.; Yoder, B.K. Mutations in Traf3ip1 reveal defects in ciliogenesis, embryonic development, and altered cell size regulation. *Dev. Biol.* **2011**, *360*, 66–76.
88. Davis, B.N.; Hilyard, A.C.; Lagna, G.; Hata, A. SMAD proteins control DROSHA-mediated microRNA maturation. *Nature* **2008**, *454*, 56–61.
89. Goizet, C.; Boukhris, A.; Mundwiller, E.; Tallaksen, C.; Forlani, S.; Toutain, A.; Carriere, N.; Paquis, V.; Depienne, C.; Durr, A.; *et al.* Complicated forms of autosomal dominant hereditary spastic paraplegia are frequent in SPG10. *Hum. Mutat.* **2008**, *30*, E376–E385.
90. Youngs, E.L.; Hellings, J.A.; Butler, M.G. ANKRD11 gene deletion in a 17-year-old male. *Clin. Dysmorphol.* **2011**, *20*, 170–171.

91. Sirmaci, A.; Spiliopoulos, M.; Brancati, F.; Powell, E.; Duman, D.; Abrams, A.; Bademci, G.; Agolini, E.; Guo, S.; Konuk, B.; *et al.* Mutations in ANKRD11 cause KBG syndrome, characterized by intellectual disability, skeletal malformations, and macrodontia. *Am. J. Hum. Genet.* **2011**, *89*, 289–294.
92. Mi, S.; Lee, X.; Hu, Y.; Ji, B.; Shao, Z.; Yang, W.; Huang, G.; Walus, L.; Rhodes, K.; Gong, B.J.; *et al.* Death receptor 6 negatively regulates oligodendrocyte survival, maturation and myelination. *Nat. Med.* **2011**, *17*, 816–821.
93. Guernsey, D.L.; Jiang, H.; Bedard, K.; Evans, S.C.; Ferguson, M.; Matsuoka, M.; Macgillivray, C.; Nightingale, M.; Perry, S.; Rideout, A.L.; *et al.* Mutation in the gene encoding ubiquitin ligase LRSAM1 in patients with Charcot-Marie-Tooth disease. *PLoS Genet.* **2010**, *6*, doi:10.1371/journal.pgen.1001081.
94. Weterman, M.A.; Sorrentino, V.; Kasher, P.R.; Jakobs, M.E.; van Engelen, B.G.; Fluiter, K.; de Wissel, M.B.; Sizarov, A.; Nurnberg, G.; Nurnberg, P.; *et al.* A frameshift mutation in LRSAM1 is responsible for a dominant hereditary polyneuropathy. *Hum. Mol. Genet.* **2012**, *21*, 358–370.
95. Burnside, R.D.; Pasion, R.; Mikhail, F.M.; Carroll, A.J.; Robin, N.H.; Youngs, E.L.; Gadi, I.K.; Keitges, E.; Jaswaney, V.L.; Papenhausen, P.R.; *et al.* Microdeletion/microduplication of proximal 15q11.2 between BP1 and BP2: A susceptibility region for neurological dysfunction including developmental and language delay. *Hum. Genet.* **2011**, *130*, 517–528.
96. Pastuszak, I.; Ketchum, C.; Hermanson, G.; Sjoberg, E.J.; Drake, R.; Elbein, A.D. GDP-L-fucose pyrophosphorylase. Purification, cDNA cloning, and properties of the enzyme. *J. Biol. Chem.* **1998**, *273*, 30165–30174.
97. Xu, C.; Li, P.P.; Cooke, R.G.; Parikh, S.V.; Wang, K.; Kennedy, J.L.; Warsh, J.J. TRPM2 variants and bipolar disorder risk: Confirmation in a family-based association study. *Bipolar Disord.* **2009**, *11*, 1–10.
98. Bueno, M.T.; Richard, S. SUMOylation negatively modulates target gene occupancy of the KDM5B, a histone lysine demethylase. *Epigenetics* **2013**, *8*, 1162–1175.
99. Catchpole, S.; Spencer-Dene, B.; Hall, D.; Santangelo, S.; Rosewell, I.; Guenatri, M.; Beatson, R.; Scibetta, A.G.; Burchell, J.M.; Taylor-Papadimitriou, J. PLU-1/JARID1B/KDM5B is required for embryonic survival and contributes to cell proliferation in the mammary gland and in ER+ breast cancer cells. *Int. J. Oncol.* **2011**, *38*, 1267–1277.
100. Carmona-Mora, P.; Walz, K. Retinoic acid induced 1, RAI1: A dosage sensitive gene related to neurobehavioral alterations including autistic behavior. *Curr. Genomics* **2010**, *11*, 607–617.
101. Van Wijk, S.J.; Timmers, H.T. The family of ubiquitin-conjugating enzymes (E2s): Deciding between life and death of proteins. *FASEB J.* **2010**, *24*, 981–993.
102. Kirchmair, R.; Hogue-Angeletti, R.; Gutierrez, J.; Fischer-Colbrie, R.; Winkler, H. Secretoneurin—A neuropeptide generated in brain, adrenal medulla and other endocrine tissues by proteolytic processing of secretogranin II (chromogranin C). *Neuroscience* **1993**, *53*, 359–365.
103. Tanji, K.; Kamitani, T.; Mori, F.; Kakita, A.; Takahashi, H.; Wakabayashi, K. TRIM9, a novel brain-specific E3 ubiquitin ligase, is repressed in the brain of Parkinson's disease and dementia with Lewy bodies. *Neurobiol. Dis.* **2010**, *38*, 210–218.

104. Takash, W.; Canizares, J.; Bonneaud, N.; Poulat, F.; Mattei, M.G.; Jay, P.; Berta, P. SOX7 transcription factor: Sequence, chromosomal localisation, expression, transactivation and interference with Wnt signalling. *Nucleic Acids Res.* **2001**, *29*, 4274–4283.
105. Agrawal, A.; Lynskey, M.T.; Hinrichs, A.; Grucza, R.; Saccone, S.F.; Krueger, R.; Neuman, R.; Howells, W.; Fisher, S.; Fox, L.; *et al.* A genome-wide association study of DSM-IV cannabis dependence. *Addict. Biol.* **2011**, *16*, 514–518.
106. Kokura, K.; Sun, L.; Bedford, M.T.; Fang, J. Methyl-H3K9-binding protein MPP8 mediates E-cadherin gene silencing and promotes tumour cell motility and invasion. *EMBO J.* **2010**, *29*, 3673–3687.
107. Bonnert, T.P.; McKernan, R.M.; Farrar, S.; le Bourdelles, B.; Heavens, R.P.; Smith, D.W.; Hewson, L.; Rigby, M.R.; Sirinathsinghji, D.J.; Brown, N.; *et al.* Theta, a novel  $\gamma$ -aminobutyric acid type a receptor subunit. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 9891–9896.
108. Kobayashi, K.; Kuroda, S.; Fukata, M.; Nakamura, T.; Nagase, T.; Nomura, N.; Matsuura, Y.; Yoshida-Kubomura, N.; Iwamatsu, A.; Kaibuchi, K. P140Sra-1 (specifically rac1-associated protein) is a novel specific target for rac1 small GTPase. *J. Biol. Chem.* **1998**, *273*, 291–295.
109. Schenck, A.; Bardoni, B.; Moro, A.; Bagni, C.; Mandel, J.L. A highly conserved protein family interacting with the fragile X mental retardation protein (FMRP) and displaying selective interactions with fmrp-related proteins FXR1P and FXR2P. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 8844–8849.
110. Yang, Y.; Muzny, D.M.; Xia, F.; Niu, Z.; Person, R.; Ding, Y.; Ward, P.; Braxton, A.; Wang, M.; Buhay, C.; *et al.* Molecular findings among patients referred for clinical whole-exome sequencing. *JAMA* **2014**, *312*, 1870–1879.

© 2015 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).