

# Genomics of Autism



**Khalid A. Fakhro**

**Abstract** Autism spectrum disorder (ASD) is a heterogeneous condition affecting >1% of all children, characterized by impaired social interactions, repetitive behavior and a widely variable spectrum of comorbidities. These comorbidities may include developmental delay, gastrointestinal problems, cardiac disorders, immune and autoimmune dysregulation, neurological manifestations (e.g., epilepsy, intellectual disability), and other clinical features. This wide phenotypic heterogeneity is difficult to predict and manifests across a wide range of ages and with a high degree of difference in severity, making disease management and prediction of a successful intervention very difficult. Recently, advances in genomics and other molecular technologies have enabled the study of ASD on a molecular level, illuminating genes and pathways whose perturbations help explain the clinical variability among patients, and whose impairments provide possible opportunities for better treatment options. In fact, there are now >1000 genes that have been linked to ASD through genetic studies of more than 10,000 patients and their families. This chapter discusses these discoveries and in the context of recent developments in genomics and bioinformatics, while also examining the trajectory of gene discovery efforts over the past few decades, as both better ascertainment and global attention have been given to this highly vulnerable patient population.

**Keywords** Autism · Genomics · Next-generation sequencing · NGS · Copy number variations · Chromosomal abnormalities · Exome sequencing · Genome sequencing · Bioinformatics

---

K. A. Fakhro (✉)  
Sidra Medicine, Doha, Qatar  
e-mail: [kfakhro@sidra.org](mailto:kfakhro@sidra.org)

## 1 Introduction

Autism spectrum disorders (ASD) are a group of highly heterogeneous disorders characterized by repetitive behaviors, impaired social interactions and a wide spectrum of neurodevelopmental and physical comorbidities. While the overall prevalence of ASD is estimated at 1 in 68 children [1], this may not represent an increase in incidence as much as it represents a widening of the scope of disorders that fit under the ASD umbrella in an era of improving clinical ascertainment. As a spectrum disorder, ASD may present as an isolated set of symptoms or with multiple comorbidities, including but not limited to intellectual disability, developmental delay, epilepsy, gastrointestinal complications, cardiac problems, immune disorders, etc. [2]. This heterogeneity is apparent even in the settings of identical genetic backgrounds (e.g., monozygotic twins discordant for co-morbidities), underscoring the complexity of understanding ASD on the molecular level.

While ASD disproportionately affects males (male: female ratio of 3.4:1), the reasons for this remain poorly understood. In fact, the search for a molecular etiology is further complicated by the interplay of both genetic and environmental factors which together contribute to pathogenesis. Importantly, high incidence despite the significant impairment of reproductive fitness means that the cause of ASD is likely different among most cases of ASD, i.e., unrelated patients will rarely share the same mutation or even the same gene. Despite that, it is clear that ASD has a major heritable component, with siblings of ASD patients usually having a one in five risk (ten-fold higher than population average) of developing ASD themselves [3]. Further, concordance between monozygotic twins ranges from 30% to 99%, and the overall heritability is estimated between 0.5 and 0.8 [4–8].

This complex landscape has made the search for and discovery of genetic factors using traditional methods very difficult in the general population, mainly due to studies being underpowered to detect causal variants in small sample sizes. As expected by the limitations imposed by previous technologies, the majority of loci identified were in the form of chromosomal abnormalities, with few individual genes identified. The subsequent introduction of high-throughput microarrays enabled the investigation of smaller chromosomal abnormalities termed copy number variations (CNVs), and study of associations between common variants and the trait of interest. Signals detected from these three approaches (linkage, karyotyping and microarrays) rarely produced single-candidate genes; usually narrowing the search space to several kilobases or megabases, in which the search for causative genes was iterative and time-consuming. Alternatively, some mutations could be found by resequencing genes known to cause similar phenotypes in model organisms in a larger patient cohort.

Subsequent technological improvements led to the advent of next-generation sequencing (NGS), which has transformed the field profoundly, allowing the discovery of different classes of variations (e.g., single nucleotide variants (SNVs) and insertions/deletions (indels)) genome-wide. This enabled making discoveries from

nuclear families in the absence of multiple affected or large pedigrees to establish linkage. The proliferation and accessibility of NGS technologies mean the bottleneck is no longer the ability to detect variants in a cost-effective manner, but the ability to amass cohorts that are large enough to capture a significant proportion of the genetic and phenotypic heterogeneity underlying ASD in the general population.

This chapter summarizes gene discovery in ASD in the pre- and post-sequencing era, explaining the significance of these discoveries, and framing them in the larger context of the potential impact that genomics will have on ASD diagnosis and care in the future.

## **2 Pre-NGS Era**

### ***2.1 Introduction***

Next-generation sequencing (NGS) refers to advancements in technology that have enabled large-scale sequencing of many DNA fragments at the same time. These advancements have allowed the interrogation of variation at many loci in the genome in parallel at reasonable speed and cost, thus increasing the efficiency of genetic research. While NGS technologies began to appear in academic environments over a decade ago [9], their proliferation and adoption into the mainstream was not until more than a decade later. Importantly, while several different technologies appeared initially to compete for adoption, it was Illumina's short-read sequencing technology that was able to capture the biggest market segment with a combination of price point, accuracy and speed. And while today the price of a single human genome is around \$1000, the price was significantly higher up until just a few years ago, rendering large-scale studies still very costly. This section covers discoveries made in autism genetics prior to the introduction to NGS to study this condition, whereas the next chapter will cover discoveries made when large-scale genomic assessment became increasingly affordable, in what is known as the post-“genomic” or post-NGS era.

### ***2.2 Linkage Studies***

Due to the paucity of multiplex or extended pedigrees with ASD, linkage approaches were not a robust approach to gene discovery in ASD. Nevertheless, numerous studies were performed (reviewed in [10]), revealing few loci in total. Notably, of these, only two were ever replicated successfully in an independent study. These include linkage to chromosome 7q35, containing the CNTNAP2 gene [11], and to chromosome 20p13, containing the four genes [12].

### 2.3 Association Studies

The development of cheap high throughput microarray genotyping technologies with higher marker density empowered a flurry of genome-wide association studies (GWAS) in a wide variety of human diseases. The discovery of markers by GWAS has two main limitations. First, they indicate loci with small effect size on the trait, sometimes increasing odds ratio by as little as 0.05 [13]. Second, they require very large sample sizes to have sufficient power to discriminate alleles between cases and controls. For ASD, large cohorts were not possible to amass for sufficient power, and therefore while GWAS were attempted over the past 10 years, few were ever replicated [12–17]. This is in contrast to studies of other neurodevelopmental conditions such as Schizophrenia for which cohorts could be amassed in the tens of thousands to discover tens of loci that replicate in independent cohorts. For ASD, only two loci have been implicated using GWAS to date, including a locus on 5p14.1 (containing the CDH9 and CDH10 genes), and another on 20p12.1 (MACRO2 gene) [15, 16]. Importantly, consistent with the genetic heterogeneity and the need for very large numbers, neither of these loci has been replicated.

### 2.4 Chromosomal Abnormalities Studies

The association of ASD with other syndromic comorbidities such as Fragile X and intellectual disability was a first indicator that chromosomal-level events could be underlying a subset of the condition. The concurrent evolution of microarray technologies introduced the ability to rapidly detect structural copy number variations in human genomes at scale. Together, karyotyping and CNV analysis have uncovered tens of chromosomal segments involved in ASD, including duplication of 15q [18], deletion of 22q11.2 [19, p. 200], deletion of 16p11.2 [20] and deletion of Xp22.3 [21]. In addition, several recurrent hotspots of de novo CNVs with ASD include duplications on 7q11.2 and deletions of 16p11.2, the latter also associated with schizophrenia [22, 23].

A key feature of CNVs is that they range widely in size from single-gene deletions to large regions encompassing tens to hundreds of genes. Consistent with multi-genic contribution to other phenotypes, patients with multiple de novo CNVs or large chromosomal abnormalities usually have more severe, syndromic phenotypes [24, p. 2], [25].

As cohort sizes grow, it has also been shown that CNV-affected genes predominantly comprise candidates from three key pathways, including neuronal signaling, synaptic function, and chromatin remodeling [26], [27, p. 201]. Together, these studies not only identify novel loci, but demonstrate that de novo CNVs are strongly associated with ASD [28] and that recurrent CNVs point to shared architecture with other diseases.

## 2.5 *Candidate Gene Resequencing Studies*

In contrast to the paucity of discovery from linkage studies, work from both other syndromes and CNV studies identified several candidate human ASD genes that could be screened by resequencing in larger cohorts. By assessing larger cohorts for mutations in these genes, the following genes were all found to harbor damaging point mutations in ASD subjects: MECP2 (Rett syndrome), TSC1 and TSC2 (tuberos sclerosis), CACNA1C (Timothy syndrome), NLGN3 and NLGN4 (X-linked mental retardation), and CNTNAP2 (7q35 deletion), SLC9A9 and BCKDK (epilepsy), etc. [29–31]. Other genes also discovered to carry rare damaging variants by resequencing include SHANK1, SHANK2, SHANK3, NRXN1 and NRXN3 [32–36].

## 2.6 *Conclusion*

In conclusion, the pre-NGS era relied mainly on candidate gene resequencing and association studies to link genes and loci to Autism. Unlike other monogenic disorders, linkage analysis was not a very successful approach to finding genes linked to Autism primarily due to the requirements of large pedigrees or multiple kindreds segregating the same locus, which are difficult to find considering the genetic heterogeneity underlying Autism and the detrimental effect it has on reproductive fitness.

# 3 *The NGS Era*

## 3.1 *Next-Generation Sequencing as a Tool to Study Genetic Disease*

Over the past decade, there have been numerous tools developed for next-generation sequencing (NGS) (Table 1). At its core, NGS may be broadly classified into two categories: whole genome sequencing (WGS), and targeted NGS. While the former is concerned with reading the entire content of an organism's genetic material, targeted NGS methods focus on selectively sequencing a group of genes ("gene panels"), usually selected based on specific selection criteria, e.g., having been identified in smaller cohorts or in animal studies, or genes within the same pathway(s) as well-established candidate disease genes. These gene panels may be customized to include any number of genomic fragments of interest, including, for example, *all* coding regions—commonly known as whole exome sequencing (WES). Typical WES experiments also capture flanking regulatory regions, enabling discovery of variants affecting splice junctions and untranslated promoter and downstream sequences [37].

**Table 1** Comparison of different sequencing technologies

Technology	Maximum reads per run	Read length per run (bp)	Accuracy (single read not consensus)	Time per run	Cost per 1 million bases (in US\$)	Advantages	Disadvantages	References
Sequencing by synthesis (SBS) Illumina	MiSeq series	2 × 150	99.9%	4–55 h	0.05 to \$0.15	Potential for high sequence yield depends on sequencer model and desired application	System can be very expensive. Requires high concentrations of DNA	[9]
	NextSeq series	2 × 150		12–30 h				
	HiSeq series	2 × 150		<1–3.5 days (HiSeq 3000/4000) 7 h–6 days (HiSeq 2500)				
Sequencing by semiconductor Thermo Fisher Scientific	HiSeq X series	200–400	98%	<3 days	\$1	Less expensive equipment	Homopolymer errors	
	Ion torrent Ion proton	125		2–4 h				
Single-molecule real-time sequencing PacBio Bioscience <sup>a</sup>	Sequel system PacBio RS II	50,000 per SMRT cell, or 500–1000 megabases	87% single-read accuracy	30 min to 6 h	\$0.13–\$0.60	Longest read length. Fast. Detects 4 mC, 5 mC, 6 mA	Moderate throughput. Equipment can be very expensive	
	MinION GridION X5 PromethION SmidgION	Dependent on library preparation, not the device, so user chooses read length (up to 200 kb reported)	~92–97% single read (up to 99.96% consensus)	Data streamed in real time. Choose 1 min to 48 h	\$500–999 per flow cell, base cost dependent on experiment	Very long reads. Portable (palm sized)	Lower throughput than other machines. Single read accuracy in 90 s	

<sup>a</sup>On 1 November 2018, Illumina entered into a purchase agreement to buy PacBio for ~\$1.2B in total

Whole genome sequencing (WGS), on the other hand, covers both WES regions as well as non-coding and inter-genic regions. It is usually faster and more uniform because it does not require target panel capture, and thus can be performed with minimal sample preparation, resulting in sequences that are evenly distributed across all chromosomes. This distribution of sequencing *coverage* means that variants can be confidently assigned at average depth of sequencing as low as 20X. Conversely, whole-exome and other panel sequencing requires target enrichment and PCR amplification, often resulting in highly variable coverage profiles with some regions (e.g., repetitive elements or GC-rich content) being missed due to the technical limitations. Another important advantage of WGS's even coverage is the ability to discover genome-wide structural variants (including copy number variants). Given the number of human disorders (including ASD) in which structural variants play a significant role, a single test that can assess both large and small genomic variation is often cited a reason to use WGS despite its slightly higher cost vis-à-vis using a combination of microarray and WES for each patient.

### ***3.2 Bioinformatics and Variant Interpretation***

One important aspect of the NGS approach is the generation of large quantities of data, often requiring sophisticated computational tools (bioinformatics) to interpret. Specifically, bioinformatics pipelines share three major steps in common, irrespective of the NGS technology used: read alignment to a reference genome, variant calling versus the reference, and variant interpretation to determine pathogenic from benign variation.

Genetic variants may belong to several different classes, including: single nucleotide variants (SNVs, including single nucleotide polymorphisms (SNPs)), multi-nucleotide variants (MNVs, including small insertions and deletions (indels)), and structural variations (SVs, including copy number variations (CNVs)). For all three variant classes, a number of statistical considerations need to be taken into account to sort out likely true positive variants from noise, including: depth of sequencing, sequencing quality, the number of times mutations are observed, and the likelihood that such a change is true rather than an artifact of sequencing [38]. Importantly, the joint steps of read alignment and variant calling may themselves introduce error into the experiment, e.g., for fragments coming from highly repetitive genomic segments [39, 40].

The most challenging aspect in bioinformatics pipeline is variant interpretation—the step where tens to hundreds of variants may require in-depth manual scrutiny to determine putative effect on disease. Robust variant interpretation requires a well-annotated reference genome (for both coding and non-coding elements) and a large number of control individuals to accurately discriminate putative disease causing variants from population-specific polymorphisms (that may rarely appear in public databases because inadequate numbers of population-matched controls are available) [41–45].

### **3.3 *NGS Suitability in Routine Clinical Care***

As NGS technologies become more widely adopted in academic hospital settings, there is a growing need to establish gold-standard pipelines to allow for genomics to enter routine clinical testing [46, 47]. While some guidelines do exist, especially for diagnostic laboratory settings, these guidelines vary widely and currently still require orthogonal validation before they are deemed actionable [46, 48]. The role of a clinical-grade pipeline is primarily to demonstrate processing and interpretation in a highly reproducible manner, thus ensuring disease management is not compromised from this approach [47]. These steps, however, are non-trivial—they would need to account for influences on data quality and sources of error, for example, sample prep using protocols, sequencing instruments and batch effects, ensuring all genes in a panel are adequately captured, errors in sequencing chemistry and noise from the sequence alignment and variant calling steps.

Further, these tasks scale in complexity with the number of samples being studied and the databases from which annotations are being drawn. Of key consideration, for example, is the large number of variant sites produced per NGS run (three to four million per genome). Amongst these, hundreds or thousands of variants would be considered variants of unknown significance (VUS) whose interpretation and relevance to health and disease is completely unknown [46, 49]. In many cases, the recruitment of parents and siblings could help with sorting through these variants, but still tens to hundreds remain “private” variants with unknown function. For NGS to be adopted in routine care, clinical platforms must deal with such cases systematically, bearing in mind not to discard these variants because they may have future value as the genome is better annotated in the academic literature. Moreover, clinical platforms should take into consideration the constantly evolving annotations of genes, e.g., >200 new genes and hundreds of variants are being linked to diseases each year [50–53], and thus variant sharing as part of consortia may mitigate the absence of variants in the publication record. Such considerations need to be taken into account when designing clinical NGS pipelines, to ensure that genetic testing of patients is accurate, reproducible and safe. Only by controlling for these factors in a statistically robust framework would it be possible to ensure reproducibility and standardization, thereby enabling precision in data interpretation in disease settings.

## **4 Successful Application of NGS to Autism Spectrum Disorder**

### **4.1 *Sample Size and Cohort Considerations***

The evolution of NGS thus enables the assessment of single families and single cases at a rate not performed before. The biggest challenge lies in discriminating rare alleles from population-specific polymorphisms, a challenge that can only be adequately addressed by sequencing a large enough number of both patients and of



ethnic/population-matched controls. This is especially important in the setting of high genetic heterogeneity, where it is unlikely to find individuals sharing mutations in the same gene, let alone the same pathogenic variant. In recent studies, for example, sample sizes of >2000 families were required to identify recurrent gene and copy number regions shared between individuals [20, 23, 27, 54].

Conversely, in settings with high consanguinity, the approach of finding recessive variants is boosted by the usual availability of affected siblings or additional cousins (in multiplex families) who share the same homozygous mutations in candidate genes. However, the identification of recessive genes causing ASD has been limited so far by the type of families studied—mostly outbred simplex families with unrelated parents. In rare cases where families with multiple affected siblings were identified, they were found to have two different de novo causative variants rather than the same recessive variant [55]. This is expected due to the high levels of genetic heterogeneity underlying ASD.

However, this presents an important opportunity for consanguineous populations attempting ASD studies, with some initial reports reporting promising results [56–59].

## 4.2 Exome and Genome Sequencing

Due to the paucity of studies in ASD families from areas of high consanguinity, recessive variants causing ASD have only been identified so far in the following genes: *AMT*, *BCKDK*, *CNTNAP2*, *PEX7*, *SLC9A9*, *SYNE1*, *VPS13B*, *PAH* and *POMGNT1* [60]. In contrast to the few recessive genes discovered, the vast majority of families studied to date have been outbred, in which single affecteds (simplex) are born to unaffected, unrelated parents. In these cases, the genetic architecture is usually driven by de novo mutations, or rare inherited variants; however, even when multiple siblings are found in the same family, they are sometimes found to harbor separate de novo variants, stressing the importance of this type of variation in ASD etiology. There are approximately 800 genes affected by de novo variants in ASD (not counting genes within de novo chromosomal abnormalities) [10]. Altogether, the contribution of de novo mutations in ASD is estimated to be between 15% and 25% [61].

In 2012, four groups published concurrent studies using exome-sequencing to identify de novo gene disrupting variants in ASD patients. Only approximately 20 of these genes were recurrently hit across the cohorts, including: *ADNP*, *ANK2*, *ARID1B*, *BCL11A*, *CACNA2D3*, *CHD8*, *CUL3*, *DSCAM*, *DYRK1A*, *GRIN2B*, *KDM5B*, *KDM6B*, *KMT2C*, *KMT2E*, *KMT5B*, *NCKAP1*, *PHF2*, *RIMS1*, *SCN2A*, *SYNGAP1*, *TBR1*, *TCF7L2*, *TNRC6B*, and *WAC* [22, 54, 62–67]. However, the majority of the other genes identified were singletons (only observed in a single patient without replication), but their potential role in ASD was supported by their impacting critical pathways in neurological development, such as cognition, synaptic formation, and regulation of transcription of brain-specific genes [54, 67, 68].

In addition to *de novo* variants affecting genes directly, more recent studies have found an enrichment of *de novo* and private disruptive mutations in DNase I hypersensitivity sites in regions close to some of the genes that have been implicated in ASD [69]. This indicates that ASD genes disruption is not only through mutations that may alter function but also those that may alter gene regulation. Notably, one recurrent theme across most studies is that *de novo* point-mutations are predominantly paternal in origin, with the rate of *de novo* mutations increasing with paternal age.

Despite advances in WES, less than 10% of known patients receive a genetic diagnosis in ASD. This is far lower than the solve rate of neurodevelopmental disorders as a whole, with diagnoses above 30% of cases. Nevertheless, the utility of WES and WGS extends beyond simple diagnostic value, as it has allowed the identification of genes underlying more complex syndromes shared with ASD. For example, *de novo* mutations in the SWI-SNF-related gene *ADNP* causes a syndromic form of ASD with unique facial dysmorphism [70], whereas mutations in the NatA complex subunit *NAA15* cause a syndromic form of ASD with multiple congenital anomalies including craniofacial, neuromuscular, and cardiac complications [71]. Such families may not have been individually identified *a priori* to share similar genetic underpinnings prior to the advent of NGS technologies, which now enable patients to be stratified more precisely based on their genetic abnormality rather than phenotypic variability.

Importantly, accurate genetic diagnosis is critical for determining potential therapeutic approaches for patients with ASD. One area where the impact has been most recognizable is in ASD related to branched-chain amino acid deficiencies, for example, branched-chain keto-acid dehydrogenase kinase deficiency, in which mutations in *BCKDK* were identified. These mutations cause loss of function of BCKDK, itself a repressor of branched-chain amino acid degradation, and therefore patients have a concurrent deficiency of BCAAs. In murine models, supplementation of knockout mice with BCAAs significantly improved their neurologic phenotypes, suggesting that patients with BCKDK mutations may benefit from dietary supplementation of BCAAs to counteract the elevated degradation caused by the genetic mutation [31]. Similarly, ASD patients with a wide variety of comorbidities (e.g., sleep disorders, seizures and metabolic and immune abnormalities) have been found to have imbalances in compounds that could easily be rectified by dietary intervention, such as folate, carnitine, cobalamin, etc. [72]. More recently, one case-control randomized trial has demonstrated that supplementation with essential fatty acids, carnitine, digestive enzymes, and a hypoallergenic diet (e.g., gluten, soy, and casein-free) all improved ASD symptoms, including non-verbal IQ and nutritional status [73]. Therefore, as more cohorts of patients continue to be evaluated at the genomic and epidemiological levels, the future of ASD research can lead to novel tools and therapies that improve stratification and clinical management of patients based on their genomic information, ushering in an era of personalized medicine for ASD.

## 5 Conclusion

Autism spectrum disorders (ASD) are a heterogeneous group of disorders characterized by clinical comorbidities and extreme genetic heterogeneity. While a lot has been achieved to understand the molecular and genetic etiology, there is still a long way to go to understand how perturbations in genes ultimately lead to an ASD phenotype. Importantly, further studies may also reveal genetic markers of the development of different physical comorbidities, which can help in patient stratification and early intervention in cases predicted to become severe. Thus, as future studies are conceived, they ought not to only focus broadly on ASD patients across the entire spectrum, but also on important concepts such as data sharing and collaborations to aid in the interpretation, and eventually treatment of ASD across the globe.

## References

1. Elsabbagh, M., et al. (2012). Global prevalence of autism and other pervasive developmental disorders. *Autism Research - Wiley Online Library*. [Online]. Retrieved January 28, 2019, from <https://onlinelibrary.wiley.com/doi/full/10.1002/aur.239>
2. Aldinger, K. A., et al. (2015). Patterns of risk for multiple co-occurring medical conditions replicate across distinct cohorts of children with autism spectrum disorder. *Autism Research - Wiley Online Library*. [Online]. Retrieved January 28, 2019, from <https://onlinelibrary.wiley.com/doi/full/10.1002/aur.1492>
3. Ozonoff, S., et al. (2011). Recurrence risk for autism spectrum disorders: A Baby Siblings Research Consortium study. *Pediatrics*, *128*(3), e488–e495.
4. Bailey, A., et al. (1995). Autism as a strongly genetic disorder: Evidence from a British twin study. *Psychological Medicine*, *25*(01), 63.
5. Rosenberg, R. E., Law, J. K., Yenokyan, G., McGready, J., Kaufmann, W. E., & Law, P. A. (2009). Characteristics and concordance of autism spectrum disorders among 277 twin pairs. *Archives of Pediatrics & Adolescent Medicine*, *163*(10), 907.
6. Hallmayer, J. (2011). Genetic heritability and shared environmental factors among twin pairs with autism. *Archives of General Psychiatry*, *68*(11), 1095.
7. Sandin, S., Lichtenstein, P., Kuja-Halkola, R., Larsson, H., Hultman, C. M., & Reichenberg, A. (2014). The familial risk of autism. *JAMA*, *311*(17), 1770.
8. Colvert, E., et al. (2015). Heritability of autism spectrum disorder in a UK population-based twin sample. *JAMA Psychiatry*, *72*(5), 415.
9. Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: Ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, *17*(6), 333–351.
10. Geschwind, D. H., & State, M. W. (2015). Gene hunting in autism spectrum disorder: On the path to precision medicine. *The Lancet Neurology*, *14*(11), 1109–1120.
11. Alarcón, M., Cantor, R. M., Liu, J., Gilliam, T. C., & Geschwind, D. H. (2002). Evidence for a language quantitative trait locus on chromosome 7q in multiplex autism families. *The American Journal of Human Genetics*, *70*(1), 60–71.
12. Weiss, L. A., et al. (2009). A genome-wide linkage and association scan reveals novel loci for autism. *Nature*, *461*(7265), 802–808.
13. Anney, R., et al. (2012). Individual common variants exert weak effects on the risk for autism spectrum disorders. *Human Molecular Genetics*, *21*(21), 4781–4792.
14. Ma, D., et al. (2009). A genome-wide association study of autism reveals a common novel risk locus at 5p141. *Annals of Human Genetics*, *73*(3), 263–273.

15. Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: A revolutionary tool for transcriptomics. *Nature Reviews. Genetics*, *10*(1), 57–63.
16. Anney, R., et al. (2010). A genome-wide scan for common alleles affecting risk for autism. *Human Molecular Genetics*, *19*(20), 4072–4082.
17. Chaste, P., et al. (2015). A genome-wide association study of autism using the Simons simplex collection: Does reducing phenotypic heterogeneity in autism increase genetic homogeneity? *Biological Psychiatry*, *77*(9), 775–784.
18. Bunday, S., Hardy, C., Vickers, S., Kilpatrick, M. W., & Corbett, J. A. (1994). Duplication of the 15q11-13 region in a patient with autism, epilepsy and ataxia. *Developmental Medicine and Child Neurology*, *36*(8), 736–742.
19. Fine, S. E., et al. (2005). Autism spectrum disorders and symptoms in children with molecularly confirmed 22q11.2 deletion syndrome. *Journal of Autism and Developmental Disorders*, *35*(4), 461–470.
20. Weiss, L. A., et al. (2008). Association between microdeletion and microduplication at 16p11.2 and autism. *New England Journal of Medicine*, *358*(7), 667–675.
21. Thomas, N. S., et al. (1999). Xp deletions associated with autism in three females. SpringerLink. [Online]. Retrieved January 28, 2019, from <https://link.springer.com/article/10.1007%2Fs004390050908>
22. Sanders, S. J., et al. (2011). Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams Syndrome Region, are strongly associated with autism. *Neuron*, *70*(5), 863–885.
23. Kumar, R. A., et al. (2008). Recurrent 16p11.2 microdeletions in autism. *Human Molecular Genetics*, *17*(4), 628–638.
24. Marshall, C. R., et al. (2008). Structural variation of chromosomes in autism spectrum disorder. *The American Journal of Human Genetics*, *82*(2), 477–488.
25. Pinto, D., et al. (2010). Functional impact of global rare copy number variation in autism spectrum disorders. *Nature*, *466*(7304), 368–372.
26. Pinto, D., et al. (2014). Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. *The American Journal of Human Genetics*, *94*(5), 677–694.
27. Sanders, S. J., et al. (2015). Insights into autism spectrum disorder genomic architecture and biology from 71 risk loci. *Neuron*, *87*(6), 1215–1233.
28. Levy, D., et al. (2011). Rare de novo and transmitted copy-number variation in autistic spectrum disorders. *Neuron*, *70*(5), 886–897.
29. Strauss, K. A., et al. (2006). Recessive symptomatic focal epilepsy and mutant contactin-associated protein-like 2. *The New England Journal of Medicine*, *354*(13), 1370–1377.
30. Morrow, E. M., et al. (2008). Identifying autism loci and genes by tracing recent shared ancestry. *Science*, *321*, 218.
31. Novarino, G., et al. (2012). Mutations in BCKD-kinase lead to a potentially treatable form of autism with epilepsy. *Science*, *338*(6105), 394–397.
32. Moessner, R., et al. (2007). Contribution of SHANK3 Mutations to Autism Spectrum Disorder. *The American Journal of Human Genetics*, *81*(6), 1289–1297.
33. Kim, H.-G., et al. (2008). Disruption of neuroligin 1 associated with autism spectrum disorder. *The American Journal of Human Genetics*, *82*(1), 199–207.
34. Berkel, S., et al. (2010). Mutations in the SHANK2 synaptic scaffolding gene in autism spectrum disorder and mental retardation. *Nature Genetics*, *42*(6), 489–491.
35. Berkel, S., et al. (2012). Inherited and de novo SHANK2 variants associated with autism spectrum disorder impair neuronal morphogenesis and physiology. *Human Molecular Genetics*, *21*(2), 344–357.
36. Vaags, A. K., et al. (2012). Rare deletions at the neuroligin 3 locus in autism spectrum disorder. *The American Journal of Human Genetics*, *90*(1), 133–141.
37. Bamshad, M. J., et al. (2011). Exome sequencing as a tool for Mendelian disease gene discovery. *Nature Reviews Genetics*, *12*(11), 745–755.
38. Van der Auwera, G. A., et al. (2013). From FastQ data to high confidence variant calls: The Genome Analysis Toolkit best practices pipeline. *Current Protocols in Bioinformatics*, *43*, 11.10.1–11.1033.

39. Pirooznia, M., et al. (2014). Validation and assessment of variant calling pipelines for next-generation sequencing. *Human Genomics*, 8(1), 14.
40. Pirooznia, M., Goes, F. S., & Zandi, P. P. (2015). Whole-genome CNV analysis: Advances in computational approaches. *Frontiers in Genetics*, 06, 138.
41. The 1000 Genomes Project Consortium, et al. (2015). A global reference for human genetic variation. *Nature*, 526(7571), 68–74.
42. The 1000 Genomes Project Consortium, et al. (2015). An integrated map of structural variation in 2,504 human genomes. *Nature*, 526(7571), 75–81.
43. Exome Aggregation Consortium, et al. (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature*, 536(7616), 285–291.
44. Durbin, R. M., et al. (2010). A map of human genome variation from population-scale sequencing. *Nature*, 467(7319), 1061–1073.
45. Fakhro, K. A., et al. (2016). The Qatar genome: A population-specific tool for precision medicine in the Middle East. *Human Genome Variation*, 3, 16016.
46. For the Working Group of the American College of Medical Genetics and Genomics Laboratory Quality Assurance Committee, et al. (2013). ACMG clinical laboratory standards for next-generation sequencing. *Genetics in Medicine*, 15(9), 733–747.
47. Moorthie, S., Hall, A., & Wright, C. F. (2013). Informatics and clinical genome sequencing: Opening the black box. *Genetics in Medicine*, 15(3), 165–171.
48. Gargis, A. S., et al. (2012). Assuring the quality of next-generation sequencing in clinical laboratory practice. *Nature Biotechnology*, 30(11), 1033–1036.
49. Green, R. C., et al. (2013). ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genetics in Medicine*, 15(7), 565–574.
50. Bamshad, M. J., et al. (2012). The Centers for Mendelian Genomics: A new large-scale initiative to identify the genes underlying rare Mendelian conditions. *American Journal of Medical Genetics Part A*, 158A(7), 1523–1525.
51. Yang, Y., et al. (2013). Clinical whole-exome sequencing for the diagnosis of Mendelian disorders. *New England Journal of Medicine*, 369(16), 1502–1511.
52. Jurgens, J., et al. (2015). Assessment of incidental findings in 232 whole-exome sequences from the Baylor–Hopkins Center for Mendelian Genomics. *Genetics in Medicine*, 17(10), 782–788.
53. On behalf of the IRDiRC Consortium Assembly, et al. (2017). The International Rare Diseases Research Consortium: Policies and guidelines to maximize impact. *European Journal of Human Genetics*, 25(12), 1293–1302.
54. O’Roak, B. J., et al. (2014). Recurrent de novo mutations implicate novel genes underlying simplex autism risk. *Nature Communications*, 5(1), 5595.
55. Yuen, R. K. C., et al. (2017). Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. *Nature Neuroscience*, 20(4), 602–611.
56. Rajab, A., et al. (2015). Recessive *DEAF1* mutation associates with autism, intellectual disability, basal ganglia dysfunction and epilepsy. *Journal of Medical Genetics*, 52(9), 607–611.
57. Al-Mubarak, B., et al. (2017). Whole exome sequencing reveals inherited and de novo variants in autism spectrum disorder: A trio study from Saudi families. *Scientific Reports*, 7(1), 5679.
58. Leblond, C. S., et al. (2019). Both rare and common genetic variants contribute to autism in the Faroe Islands. *NPJ Genomic Medicine*, 4, 1.
59. Jamra, R. (2018). Genetics of autosomal recessive intellectual disability. *Medizinische Genetik*, 30(3), 323–327.
60. Yu, T. W., et al. (2013). Using whole-exome sequencing to identify inherited causes of autism. *Neuron*, 77(2), 259–273.
61. Ronemus, M., Iossifov, I., Levy, D., & Wigler, M. (2014). The role of de novo mutations in the genetics of autism spectrum disorders. *Nature Reviews Genetics*, 15(2), 133–141.
62. Iossifov, I., et al. (2012). De novo gene disruptions in children on the autistic spectrum. *Neuron*, 74(2), 285–299.
63. Neale, B. M., et al. (2012). Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature*, 485(7397), 242–245.

64. Sanders, S. J., et al. (2012). De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature*, 485(7397), 237–241.
65. Willsey, A. J., et al. (2013). Coexpression networks implicate human Midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell*, 155(5), 997–1007.
66. Dong, S., et al. (2014). De novo insertions and deletions of predominantly paternal origin are associated with autism spectrum disorder. *Cell Reports*, 9(1), 16–23.
67. O’Roak, B. J., et al. (2012). Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science*, 338(6114), 1619–1622.
68. The DDD Study, et al. (2014). Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature*, 515(7526), 209–215.
69. Turner, T. N., et al. (2016). Genome sequencing of autism-affected families reveals disruption of putative noncoding regulatory DNA. *The American Journal of Human Genetics*, 98(1), 58–74.
70. Helsmoortel, C., et al. (2014). A SWI/SNF-related autism syndrome caused by de novo mutations in ADNP. *Nature Genetics*, 46(4), 380–384.
71. Cheng, H., et al. (2018). Truncating variants in NAA15 are associated with variable levels of intellectual disability, autism spectrum disorder, and congenital anomalies. *American Journal of Human Genetics*, 102(5), 985–994.
72. Frye, R. E., & Rossignol, D. A. (2016). Identification and treatment of pathophysiological comorbidities of autism spectrum disorder to achieve optimal outcomes. *Clinical Medicine Insights. Pediatrics*, 10, 43–56.
73. Adams, J. B., et al. (2018). Comprehensive nutritional and dietary intervention for autism spectrum disorder—a randomized, controlled 12-month trial. *Nutrients*, 10(3), pii: E369.