

Review

Autism spectrum disorder genomics: The progress and potential of genomic technologies

Fiana Ní Ghrálaigh^{a,b}, Louise Gallagher^a, Lorna M. Lopez^{a,b,*}

^a Department of Psychiatry, Trinity College Dublin, Ireland

^b Department of Biology, Maynooth University, Ireland

ARTICLE INFO

Keywords:

Autism
Genomics
Rare
Variation
Whole-genome
Sequencing
ASD

ABSTRACT

Genomic technologies have accelerated research progress in autism spectrum disorder (ASD) genomics and promises to further transform our understanding of the genetic basis of this neurodevelopmental disorder. Here we review the current evidence for the genetic basis of ASD, present the progress of large-scale studies to date and highlight the potential of genomic technologies. In particular, we discuss evidence for the importance of identifying rare genetic variants in family-based studies. Genomics is a key feature of future healthcare and our review illustrates its potential to improve our biological understanding of neurodevelopmental disorders, and to ultimately aid ASD diagnosis, inform medical decision making and establish genomics as central to personalised medicine.

1. The genetic basis of ASD

Autism Spectrum Disorder (ASD) is a prevalent neurodevelopmental disorder occurring in around 1% of individuals in a population [1]. The condition manifests as restrictive repetitive behaviours and social communication deficits across a phenotypic spectrum [2]. ASD is a highly heritable complex trait. The heritability of ASD measures the genomic variation contributing to the phenotype and in ASD has been estimated at ~80–90% [3,4]. The genetic risk of ASD is contributed to by both rare and common genetic variants, and as yet the majority of the genetic risk remains unexplained [5]. Rare variants refer to those occurring at less than 5% of the population and very rare variants occur at a minor allele frequency of less than 1%. Common genetics variants typically refer to genetic variants with a minor allele frequency of greater than 5%. Rare variants, particularly those occurring *de novo*, have the potential to occur at higher effect sizes than common variants. The larger effect size of rare variants is in line with the hypothesis that variants of a higher effect sizes have a more detrimental effect on brain development resulting in the early-life manifestation of the autistic phenotype, when compared to neuropsychiatric disorders most commonly arising later in life, such as schizophrenia and psychosis.

In this review, we aim to inform the reader on state-of-the-art ASD genomics research. Our focus is on the application of genomic sequencing technologies to search for these genetic variants in extensive sample collections that have transformed our understanding of ASD

genomics. We review cutting-edge research that use genomic sequencing methods, bioinformatic processing and clinical implementation for improved diagnosis and medical decision-making in ASD and other neurodevelopmental disorders. We explain the value of genomic sequencing technologies and highlight what they can achieve for neurodevelopmental and neuropsychiatric disorders.

2. Sequencing technologies have advanced the identification of rare variants

Genomic sequencing, specifically whole exome sequencing and whole genome sequencing, has transformed variant discovery. These technologies give the opportunity for more widespread and in-depth genomic analysis than older techniques, such as microarray studies and candidate gene studies, have allowed. Table 1 lists the next-generation sequencing technologies that can identify single nucleotide variants and insertion-deletion variants, as well as larger genomic hits, including structural or copy number variants, across the allele frequency spectrum. In the past decade, sequencing technologies have stretched from covering select points across to genome to cover up to 100%, when sequenced at high coverage with *de novo* assembly (Table 1) [6]. Higher coverage whole genome sequencing results in more precise variant calls across the coding and non-coding regions of the genome.

These advances in genomic technologies and decreasing costs have enabled large sequencing cohorts (Table 2), allowing key strides to be

* Corresponding author at: Department of Psychiatry, Trinity Centre for Health Sciences, St. James Hospital, Dublin 8, Ireland.
E-mail address: lorna.lopez@tcd.ie (L.M. Lopez).

<https://doi.org/10.1016/j.ygeno.2020.09.022>

Received 17 June 2020; Received in revised form 1 September 2020; Accepted 8 September 2020

Available online 15 September 2020

0888-7543/ © 2020 Elsevier Inc. All rights reserved.

Table 1
Genomic technologies compared.

	Exome sequencing		Whole genome sequencing	
	Clinical exome sequencing	Whole exome sequencing	Short-read	Long-read
% Genome covered	~0.5%	~1%	~90%	Potential for up to 100%
Types of variant detected	SNVs Indels CNVs (limited)	SNVs Indels CNVs (limited) SVs (limited) Mitochondrial	SNVs Indels CNVs SVs Mitochondrial Repeat expansions (including tandem repeats [69,70])	SNVs Indels CNVs SVs Mitochondrial Repeat expansions Complex SVs Haplotype phased variants Methylation Not yet available
Diagnostic yield in ASD	Limited application	31% [62]	42.4% [42]	Not yet available
Cost estimate	€37.19 ^a	€79.33 ^b	€1239.50 ^c	€918 ^d

Outlined are four key sequencing technologies with potential for use to identify rare ASD genetic variants. Note that these costs are estimates and do not include library preparation costs, barcodes, access fees, labour, VAT, service provider, data processing and data storage and other associated sequencing costs.

Estimates a, b and c, are based on sequencing with Illumina NovaSeq S4 flowcell (2 × 150) up to 3000Gb/flowcell.

Acronyms; SNV single nucleotide variant, Indel insertion deletion, CNV copy number variant, SV structural variant.

^a SOPHiA GENETICS Clinical Exome Solution (12 Mb covering ~4500 genes (2.5Gb/sample/800 samples/flowcell)).

^b Illumina Nextera Rapid Capture Exome (37 Mb (8Gb/sample/375 samples/flowcell)).

^c WGS (120Gb/sample/24 sample/flowcell).

^d Oxford Nanopore Technologies (60 ×; 1 sample/flow cell/180GB) Sequencing metrics: <https://nanoporetech.com/accuracy>.

made in the field of ASD genomics. Large-scale analyses of these cohorts have identified hundreds of ASD-associated genetic variants across the genome. For example, discovery of rare variants, particularly rare CNVs, affecting *SHANK3* and *NRXN1* among other genes, implicated synaptic transmission and plasticity in ASD neurobiology [7]. Extending beyond variant discovery, combining rare variant analysis with single-cell investigation in the developing human cortex showed enriched expression of particular ASD-associated genes in maturing and mature excitatory and inhibitory neurons from mid-fetal development, and helped to validate the role of these genes in neuronal communication and regulation of gene expression [8]. Impactful findings such as these, suggest great potential for advancing our understanding of ASD neurobiology through rare variant discovery.

3. Common genetic variants have been challenging to associate with autism

The search for common genetic variants has been less successful than that in more typically adult-onset neuropsychiatric disorders, in particular schizophrenia (~7% of variance on the liability scale) [9] and bipolar disorder (~2.5% of variance on the liability scale) [10–12]. The largest study to date investigating common genetic variants in ASD, using genome-wide genotyping, provides evidence for statistically significant association of the first common risk variants with ASD. A Genome Wide Association Study (GWAS) was carried out on 18,381 ASD cases and 27,969 controls. While this sample size is large in terms of ASD, it is smaller than that of other traits such as schizophrenia with 36,989 cases or bipolar disorder with 20,352 cases [9,10]. Five loci showed significant association with ASD alone and seven further loci were identified upon analysis of schizophrenia, depression and educational attainment together [13]. Polygenic risk, measured by a polygenic risk score (PRS), is the combined impact of common variants on the probability of a phenotype. In ASD this explains just 2.5% of the observed variance in risk [13]. The lower yield of common variant loci in ASD may be because of a greater relative contribution of rare genetic variants than common variants in the genetic architecture of ASD [14]. However, the current smaller sample sizes in GWAS of ASD fail to validate this hypothesis.

4. Heterogeneity in the genetic architecture of ASD

ASD displays a high level of heterogeneity across a phenotypic spectrum, both between individuals and within the same individual throughout the lifespan. It is estimated that around 10% of individuals affected with ASD have a syndromal form of the condition, for which each single ASD risk gene accounts for at most 1% of overall cases on average [15]. Rare disorders often manifest with an underlying autistic phenotype [16]. These syndromes are frequently caused by highly penetrant variants in single genes, such as Fragile X syndrome, MIM:30024 (*FMR1*), and Tuberous Sclerosis Complex, MIM: 613254 (*TSC2*) (reviewed in Betancur, [17]). These syndromal forms of ASD are frequently associated with intellectual disability and developmental delay, suggesting that ASD may only form part of the overall behavioural phenotype of the syndrome.

ASD cases that do not fall into clinically defined syndromes appear to have more complex genetic architecture and various models of risk have been suggested to encompass this. The polygenic model, strongly supported in schizophrenia [18], proposes that multiple loci, each contributing a small effect, accumulate to surpass a threshold of disease liability. In contrast, Boyle et al. proposed the omnigenic model [19,20]. This model suggests that all genes expressed in disease-relevant cells have the ability to influence pathogenesis, through their interference with the expression of “core genes”. In that, it may be hypothesised that most of the heritability of ASD could be explained by the effect of variation on genes outside of the core ASD pathways.

Understanding gene regulation is critical to parsing out the relative contribution of common and rare variants to ASD heritability. Whichever model is most appropriate in describing its architecture, it is clear that rare genetic variants are crucial to understanding ASD.

Further to heterogeneity in the genetic architecture among ASD cases, there is heterogeneity, both genetically and clinically, between males and females. Males are more frequently affected with ASD than females [21]. Although factors such as hormonal sex differences, sex-specific epigenetic factors and genetic factors related to sex chromosomes have been hypothesised to play a role in this bias, the biological basis remains unclear. A large-scale family study interrogating *de novo* variants in ASD reinforces the importance of evaluation of the X chromosome, identifying 5 of 7 genes replicated in the study are located on the X chromosome [22]. Together with the evidence of sex biases of autosomal genes, this study highlights the potential for genomic studies

Table 2
Key ASD genomics cohorts.

Cohort	Size of cohort	study design	Dataset	Reference
Australian National Autism Consortium Autism Genetic Resource Exchange (AGRE)	48 cases and 80 parent controls > 1700 families	Simplex & multiplex Simplex & multiplex	WES Genome-wide genotyping	71 https://www.autismspeaks.org/agre
Autism Sequencing Consortium (ASC) Deciphering Developmental Disorders (DDD)	12,772 individuals 12,000 individuals and their parents	Case control, simplex Simplex	WES Genotyping & WES	41 8,72 https://www.addtk.org/ 25,47
IHART iPsych Danish Cohort	2308 individuals (from 493 AGRE families) 16,146 cases (Genotyping) and 4811 cases (WES)	Quad & multiplex Case control	WGS Genome-wide genotyping (PsychChip array from Illumina) & WES WGS	33, 8,74,75
MSSNG	11,312 individuals (4258 families)	Simplex & multiplex	WGS	https://research.mss.ng/ 31,34,42,43,76
Simons Foundation Powering Autism Research for Knowledge (SPARK)	27,615 individuals (genotyping & WES) and 400 quad families (WGS)	Simplex & multiplex	Genome-wide genotyping (Illumina InfiniumCoreExome-24), WES & WGS	https://sparkforautism.org/ 40
Simons Simplex Collection (SSC)	8975 individuals (WGS) 2517 families (WES) and 10,220 individuals (Genotyping)	Quartet (phase 1–3) & Trio/ Incomplete family data (Phase 4)	Genome-wide genotyping, WES & WGS	https://www.sfrari.org/resource/simons-simplex-collection 8,32,34,77–80
The Autism Genome Project ^a , consortium including TASC and AGRE samples	7917 individuals (1492 families)	Simplex & multiplex	Genome-wide genotyping (10 K SNP array and 400 microsatellite marker panel)	37,81
The Autism Simplex Collection (TASC) The Psychiatric Genetics Consortium	5444 individuals (1719 families) 18,381 cases	Simplex Case control	Genome-wide genotyping (Illumina 1 M SNP) & WES Genome-wide genotyping	26,82 13

Featured in the table are large-scale ASD cohorts used in genomic studies to date. Note that there is significant overlap of samples between these cohorts, for example, the MSSNG cohort includes samples from both AGRE and TASC. These details are subject to frequent update. Reference refers to the original research article/website linked to the cohort and research studies cited in this review that analyse these cohorts.

to elucidate this phenomenon.

5. Rare variants disrupt gene function, dosage and regulation in ASD

Current whole genome and exome sequencing technologies enable investigation of most genomic variant classes (Table 1). The consequences of such variants in the genome occur to varying effects with different degrees of penetrance, as outlined below.

5.1. Gene disruption

Gene disruption refers to the disturbance of gene expression and the impact of variation on overall gene function. The consequence of a genetic variant can be detrimental to gene function or can have little effect depending on the variant in question and the overall genome environment (Fig. 1). Genes disrupted in ASD often include those related to brain development, post-synaptic density, nerve impulse and neuron projection [23]. Much focus lies on the importance of loss of function variants and damaging missense variants in the evaluation of genetic variation on ASD. In particular, variants impacting evolutionarily conserved genes to the detriment of crucial cellular processes.

Another mechanism of gene disruption is gene rearrangement, encompassing translocations, inversions and large-scale insertions and deletions. Although varying between studies, the estimated rate of large variants in ASD is approximately 5–10% [24]. A recent study implicates rare retro-transposition derived disruption in neurodevelopmental disorders through trio-based exome sequencing analysis from the Deciphering Developmental Disorders (DDD) cohort. This mechanism of disruption is an avenue for pathogenesis which has been largely unexplored in neurodevelopmental disorders to date [25].

5.2. Gene dosage

Gene dosage refers to the number of copies of a given gene that are present in the genome of an individual. Dosage has been found to play a substantial role in ASD pathogenesis, as demonstrated through CNV analysis, *i.e.* analysis of duplication or deletion variants of > 1Kb [26]. In 2004, two groups independently identified that large scale CNVs were often overlapping with genic regions [27,28]. The influence of these CNVs means either an increase or depletion in activity of the contained genes with potential for damaging functional consequences. A comprehensive analysis identified clinically relevant CNVs in 10.5% of neurodevelopmental disorder cases investigated, with 11.4% in ASD cases. Importantly many of the CNVs identified were found to occur across multiple neurodevelopmental disorders [29].

5.3. Gene regulation

As a complex trait, non-coding variants, particularly variants affecting gene regulation are likely to influence ASD [30]. Advances in whole genome sequencing and bioinformatic tools are enabling studies of non-coding regions of the genome. Yuen et al. estimated that non-coding and genic non-coding *de novo* variants account for 15.6% and 22.5% respectively, of predicted damaging *de novo* variants in ASD cases. Non-coding elements, *e.g.* untranslated regions, regulatory sequences involved in exon skipping and DNase hypersensitivity regions were most enriched for *de novo* variants [31]. The first study significantly associating genome-wide non-coding variants with ASD shows convergence in the pathways and processes disrupted by both coding and non-coding variants in ASD, specifically in synaptic transmission and neuronal development [32]. Ruzzo et al. also provided evidence that non-coding variants impact neurobiology in ASD, reporting a recurrent 2.5 KB deletion within the promoter of *DLG2*, a gene associated with cognition and learning in mice and human [33].

Preferential transmission of structural non-coding variants has been

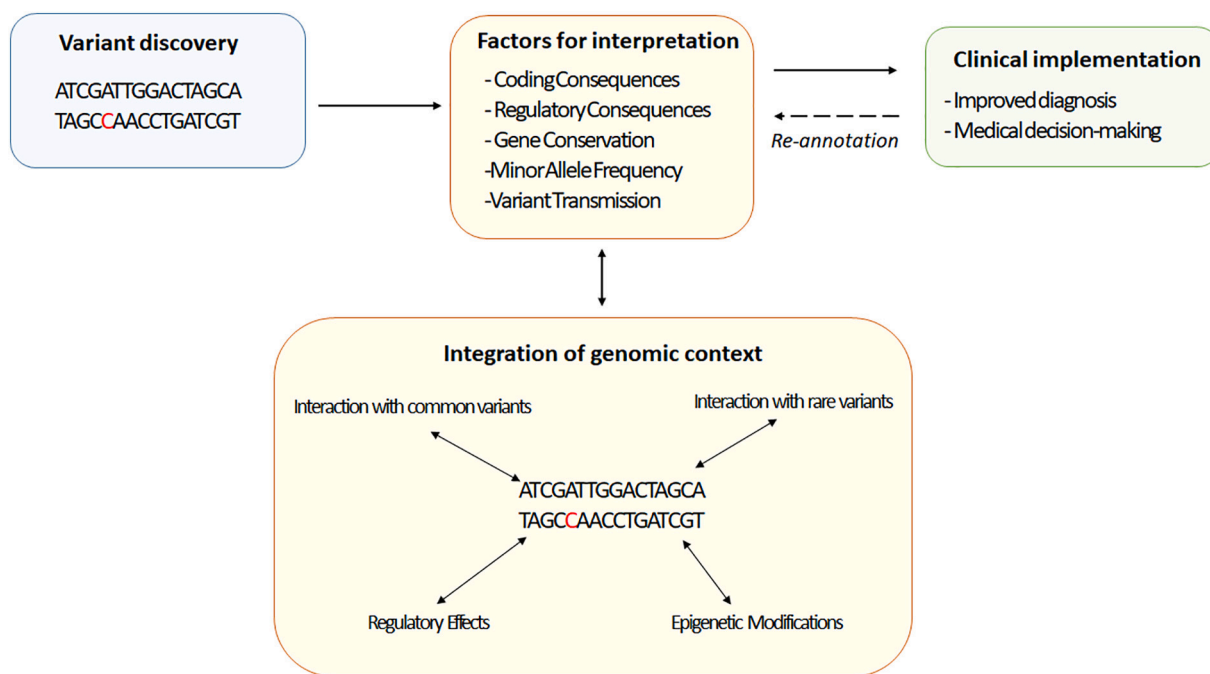


Fig. 1. Pathway from sequencing to clinical implementation. Outlined are the main stages of ASD gene discovery; from variant discovery (blue), through genomic data analysis (yellow), to accurate translation for meaningful diagnosis (green). *Re-annotation* refers to regular re-analysis of genetic diagnosis, as additional variants reach significant association with ASD. The variant highlighted in red, here a single nucleotide variant, represents any variant type detectable through application of genomic technologies (Table 1). Epigenetic modifications include methylation changes, histone modification or microRNA dysregulations (reviewed Eshraghi et al. 2018) [83]. Research is ongoing to integrate genomic variants with other variation within an individual's genome, as described by McGuire et al. (2020) [84]. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

reported in ASD, specifically the transmission of cis-regulatory elements from father to affected rather than to unaffected offspring [34]. These findings are suggestive that not only are rare inherited non-coding variants increasing risk to ASD, but also indicate a parent-of-origin effect from this non-coding variant class, highlighting a key benefit to the use of a family-based study design in studies of ASD.

6. Family-based studies are key to rare variant analysis in ASD

Family-based studies, previously the foundation of disease gene discovery, are re-emerging as an effective tool to identify potentially pathogenic variants in neuropsychiatric disorders, including ASD [35]. Family-based designs facilitate the analysis of parent to offspring variant transmission. These study designs take the form of i) simplex families (trios); parents and their affected child, ii) multiplex families; parents with more than one affected child, and iii) more complex extended pedigrees with multiple affected individuals. By design, trio studies such as those investigating the MSSNG cohort (Table 2), have been particularly key to uncovering the enrichment of *de novo* variants in cases by comparing rates of *de novo* variants in affected offspring with their unaffected respective siblings [31].

Family-based study designs also enable analyses of parent-of-origin effects that are not possible in case-control design. Furthermore, the presence of matched unaffected siblings in these studies, gives a background level of genetic variation that can be used to distinguish between disease relevant variants and those that are unrelated, such as population-specific background variation or biases introduced in sequencing. A number of large-scale genomic investigations of ASD apply a family-based approach, including the Simons Simplex Collection (Simplex), Autism Genetic Research Exchange (Simplex and Multiplex) and The Autism Genome Project (Simplex and Multiplex) (Table 2).

7. Multiplex and simplex cases of ASD show different genetic architectures

Family structure plays a major role in the types of putative variants expected to be causative of a given ASD proband. Earlier CNV studies in ASD provided some evidence of differences in genetic architecture between simplex and multiplex families [36]. These differences are centred on the contribution of *de novo* and inherited variants to ASD susceptibility.

7.1. De novo variants

A lower rate of *de novo* variation is seen in multiplex families compared to simplex families, as expected by study design. Sebat et al. reported *de novo* copy number variants in 10% of simplex cases and 3% of cases from multiplex families in their cohort [36]. Similarly Ruzzo et al. give evidence for depletion of rare *de novo* ASD risk variants in multiplex families [33]. While, this is observed across multiple studies, the difference between multiplex and simplex family structures is not consistently evident. In their CNV analyses, Pinto et al. did not report such differences [37]. A limitation to these analyses, such as analyses involving the Autism Genome Project cohort (Table 2), arises from challenges in reporting of simplex/multiplex status, *i.e.* identifying a family as a true simplex, or as a family for which just one offspring was investigated.

7.2. Inherited variants

Consistent with the enrichment of *de novo* variants in simplex cases of ASD, there is a depletion of inherited variants associated with ASD in these spontaneous cases [36,38]. Klei et al. estimate narrow sense heritability to exceed 60% for ASD cases in multiplex families but estimate just 40% of narrow sense heritability for simplex families [39]. This means that 60% of phenotypic variance may be attributed to

additive genetic variance in individuals of multiplex families. As in comparison of *de novo* variant enrichment of simplex and multiplex families, this effect is not reported consistently across analyses.

Interestingly, the same putative variant may not be found in all affected individuals within a multiplex family as highlighted recently [40]. This study reports a maternally inherited 15q11.2 deletion in an affected male child and no paternally inherited putative variants from an affected father. Other studies have identified non-sharing of CNVs [41] and single nucleotide variants (SNVs) in members of multiply affected families. In the latter study the two affected siblings did not harbour the same rare risk variant in more than half of the multiplex families studied [42]. Similarly, pathogenically significant CNVs have been identified that are transmitted to an ASD proband from an unaffected parent, and shared with an unaffected sibling [43], adding to evidence for asymptomatic carriers of neurodevelopmental disorder CNVs.

Family studies in epidemiological cohorts from isolated populations have also confirmed that both rare and common genetic variants contribute to the susceptibility to ASD. A study on the Faro Island genetic isolate, affirms the importance of both common and rare variants in ASD susceptibility [44]. This study identifies in a subset of individuals in the cohort carrying rare deleterious variants in genes known already associated with ASD and in this same cohort, common genetic variants were also associated.

Given these two mechanisms of genetic variation, *de novo* and inherited in ASD, genomic sequencing studies in families with multiple affected individuals offers greater opportunity to understand the relative contribution of inherited and *de novo* variation in the genetic architecture of ASD.

8. Establishing putative ASD variants faces many challenges

Heterogeneity in ASD diagnoses is a major challenge facing genomic sequencing studies in ASD. In particular, diagnosis of ASD in the presence of intellectual disability. Diagnostic procedures are found to differ between that used in a clinical and research setting. For a comprehensive discussion on these challenges refer to Schaaf et al. [45].

The greatest challenge in analysis of large-scale genomic data is in the establishment of pipelines for data interpretation. Interpretation of putative variants is complicated by a wide variety of technical factors, such as sequence coverage, variant validation, consistency in sequencing platforms and variant calling and filtering techniques. Robust clinical diagnoses and rich phenotyping increase confidence in variant association [46]. A variant that has been associated with ASD and has substantial evidence supporting its validity will be interrogated for its biological role (Fig. 1).

Variants associated with ASD disrupt a wide variety of pathways and biological processes [7]. Identifying pathways and processes showing an increased mutational burden enables the isolation of cellular processes and pathways disrupted in ASD. Gene-lists are often compiled listing genes involved in a given process [42]. These lists are useful in establishing the process which a putative variant may be disrupting, and such gene lists are often consulted for membership when investigating the impact of a variant [40].

The establishment and maintenance of collective databases, such as SFARI gene [23], Developmental Disorders Genotype-to-Phenotype database (DDG2P) [47] and ClinVar [48], that are openly shared among researchers give hope for the development of variant specific disease models which will expectedly lead to a greater understanding of ASD pathology. Consistent re-analysis of pathogenicity is key to gaining maximum insight from available genomic data, as proven fruitful in the re-annotation of developmental and epileptic encephalopathies genes [49] (Fig. 1). A key stride in the development of an ASD gene list comes from Schaaf et al. in their proposal to adapt the ClinGen curation framework to ASD [45]. Development of a high-confidence gene list for ASD would have great use in genomic investigation, specifically in the

development of targeted gene panels and a ‘clinical exome’. Without a consensus gene list in ASD, attempts to develop such genome analysis strategies have limited application (Table 1).

Advances in long-read sequencing technologies hold the potential for sequencing of “dark gene regions”, genomic regions inaccessible through next-generation sequencing. With high coverage and *de novo* assembly, Nanopore technologies have potential to sequence up to 100% of the genome (Table 1), with the greatest level of ‘recovered’ genes when compared with other genomic technologies, including the recovery of genes associated with ASD [50]. This technology, to our knowledge, has yet to be applied to ASD cohorts, aside from use in variant validation [34]. Long-read sequencing will enable discovery of genetic variants which have thus far been largely under-explored in ASD, such as repeat expansions, haplotype phased variants and methylation changes. Repeat expansion variants have already been associated with ASD, most notably the *FMR1* repeat expansion associated with Fragile X syndrome (MIM: 30024). As shown in an early haplotype mapping study, identification of haplotypes can succeed in identifying loci involved in ASD susceptibility [51]. Even more relevant perhaps, long-read sequencing enables the detection of CNVs and rearrangement events without the need for bioinformatic re-assembly and alignment of short reads.

9. Putting ASD in the context of other neuropsychiatric disorders

Whole genome sequencing has potential to investigate some of the major questions remaining unanswered in ASD genomics, including investigation of the overlap of ASD with other neurodevelopmental and neuropsychiatric disorders, both clinically and genetically. As highlighted in a review from Lord et al., elucidation of the genetic overlap of ASD with other neuropsychiatric disorders is needed [52]. Clinically, ASD frequently occurs co-morbidly with other neuropsychiatric disorders, in particular attention-deficit hyperactivity disorder (28%), anxiety disorders (13%) and mood disorders (11%) [53].

At the systems-level there is substantial evidence of genetic overlap between ASD and neurodevelopmental and neuropsychiatric disorders [54]. There is overlap in the genes associated with ASD and those associated with other neuropsychiatric disorders, such as schizophrenia and bipolar disorder [55–57]. This has been demonstrated strongly in a large-scale meta-analysis of eight European psychiatric cohorts identifying 109 pleiotropic loci [58]. The genetic overlap of ASD with other disorders is also evident at the variant level with *de novo* variation in ASD shared with intellectual disabilities [8] and shared with epilepsy [59].

10. Next-generation sequencing technologies improve diagnostic yield

There is a demand for clinical genetic testing in ASD [60]. Clinical CNV detection has already been translated widely, advancing the clinical genetics understanding of the condition. This translation crystallised some of the issues that will emerge with widespread translation of genomic technologies; namely clinical interpretation, relative contribution of inherited variants and particularly variant specificity to ASD. Currently no gene, which when disrupted by a pathogenic variant, has been found to confer risk to ASD without conferring risk to intellectual disability or other neurodevelopmental disorders. In the absence of appropriate study design and explicit, robust diagnoses, there is insufficient evidence to assign meaningful specificity of gene involvement in ASD [61].

Genomic technologies, given the greater proportion of the genome covered, have the potential to transform the clinical genetic understanding of the condition. This is illustrated by the increase in diagnostic yield with genomic technologies. Diagnostic yield refers to the number of cases where a putative genetic variant associated with the condition is identified in a cohort. This can be interpreted as a measure

of the utility of the technique and analysis strategy for the condition.

A recent meta-analysis scoping review states that exome sequencing is a first-tier clinical diagnostic test for individuals with neurodevelopmental disorders, defined in this study as developmental delay, intellectual disability and/or ASD [62]. The diagnostic yield for whole exome sequencing overall from these meta-analyses is 36%, surpassing the estimated 15–20% diagnostic yield of candidate gene arrays.

Using whole exome sequencing technologies, Feliciano et al. in the SPARK pilot, report a returnable genetic result in 10.4% of their cohorts affected offspring [40]. Importantly, in individuals with more complex phenotypes, such as ASD with seizures or co-morbid intellectual disability, they report a higher diagnostic yield than overall (27% and 20% respectively). This finding is consistent with other studies [62,63]. The SPARK study also reports a higher diagnostic yield in cases from multiplex families (15.2%) than simplex families (10.1%) [40].

Yuen et al. find a diagnostic yield of ASD relevant variants using whole genome sequencing to be 42.4% in their cohort of 85 multiplex families of ASD. This mirrors the diagnostic yield estimated in intellectual disability using the same sequencing platform [42,64]. The increased diagnostic yield using whole genome sequencing highlights the great potential for use of the technology in families with ASD. This estimate can be expected to increase further with developments in variant interpretation strategies and increases in sample sizes, giving more power to investigations of common variants and variants in the non-coding regions of the genome.

The clinical utility of whole genome sequencing holds great promise; however, this sequencing approach also faces major challenges. These include the need for large-cohort analyses and the failure to replicate genomic findings. One example is the report of the enrichment of *de novo* and private disruptive mutations within fetal CNS DNase I hypersensitive sites within 50 kb of genes that have been previously associated with autism risk [65] that later did not replicate despite a larger sample size [66]. Furthermore, we face limitations to the current capacity to interpret variants in the non-coding genome, as discussed by Lee & Gleeson [67]. Notwithstanding these challenges, the decrease in sequencing costs (Table 1) and the increase in sample sizes under investigation, together with the greater understanding of family inheritance will continue to give a more precise estimate of the diagnostic yield in ASD. The return of genetic results, alongside current behavioural diagnoses, may be used to improve therapeutic avenues in the future. Genetic diagnoses may also be used to inform family planning on a family-by-family basis as illustrated by a recent family study showing the CNV findings, which would have been pre-symptomatically predictive of ASD or atypical development in 7% (11 of 157) of families analysed [68].

11. Conclusion

Whole genome sequencing is the most effective technology to improve our biological understanding of neurodevelopmental disorders. With near full coverage of the human genome, coupled with the increase in sample sizes and the development of cutting-edge analytical methods, we now have the potential to identify more variants across the genome, in particular more rare pathogenic genetic variants. The detection of rare variants by genomic technologies will improve our understanding of the genetic architecture of ASD and other neurodevelopmental and neuropsychiatric disorders. With advances in biological interpretation enabling delivery of genetic discovery into clinical translation, genomic technologies will become an achievable step towards personalised family medicine, ultimately aiding ASD diagnosis and informing medical decision-making.

Declaration of Competing Interest

None.

Acknowledgements

Thank you very much to Dr. Elaine Kenny, TrinSeq–Genomics Core Facility, Dublin, Ireland for estimating Illumina sequencing costs. This publication has emanated from research supported in part by a Grant from Science Foundation Ireland under Grant No. 15/SIRG/3324.

References

- [1] G. Baird, et al., Prevalence of disorders of the autism spectrum in a population cohort of children in South Thames: the special needs and autism project (SNAP), *Lancet* 368 (2006) 210–215.
- [2] American Psychiatric Association, *Diagnostic and statistical manual of mental disorders*, 5th ed., (2013).
- [3] S. Sandin, et al., The heritability of autism spectrum disorder, *JAMA* 318 (2017) 1182.
- [4] B. Tick, P. Bolton, F. Happé, M. Rutter, F. Rijdsdijk, Heritability of autism spectrum disorders: a meta-analysis of twin studies, *J. Child Psychol. Psychiatry* 57 (2016) 585–595.
- [5] T. Gaugler, et al., Most genetic risk for autism resides with common variation, *Nat. Genet.* 46 (2014) 881–885.
- [6] K.H. Miga, et al., Telomere-to-telomere assembly of a complete human X chromosome, *Nature* (2020) 1–9, <https://doi.org/10.1038/s41586-020-2547-7>.
- [7] S. De Rubeis, et al., Synaptic, transcriptional and chromatin genes disrupted in autism, *Nature* 515 (2014) 209–215.
- [8] F.K. Satterstrom, et al., Large-scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism, *Cell* 180 (2020) 568–584.e23.
- [9] Schizophrenia Working Group of the Psychiatric Genomics Consortium, S. W. G. of the P. G, et al., Biological insights from 108 schizophrenia-associated genetic loci, *Nature* 511 (2014) 421–427.
- [10] E.A. Stahl, et al., Genome-wide association study identifies 30 loci associated with bipolar disorder, *Nat. Genet.* 51 (2019) 793–803.
- [11] B. Creese, et al., Examining the association between genetic liability for schizophrenia and psychotic symptoms in Alzheimer's disease, *Transl. Psychiatry* 9 (2019).
- [12] Psychiatric GWAS Consortium Bipolar Disorder Working Group, P, et al., Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4, *Nat. Genet.* 43 (2011) 977–983.
- [13] J. Grove, et al., Identification of common genetic risk variants for autism spectrum disorder, *Nat. Genet.* 51 (2019) 431–444.
- [14] J.A.S. Vorstman, et al., No evidence that common genetic risk variation is shared between schizophrenia and autism, *Am. J. Med. Genet. Part B Neuropsychiatr. Genet.* 162 (2013) 55–60.
- [15] B.S. Abrahams, D.H. Geschwind, Advances in autism genetics: on the threshold of a new neurobiology, *Nat. Rev. Genet.* 9 (2008) 341–355.
- [16] MENDELIAN.CO, Autism Spectrum Disorders: Definition and Top 150+ Rare Diseases Related to Them, MENDELIAN.CO., 2019 Available at: <https://www.mendelian.co/autism-spectrum-disorders-150-rare-diseases-related> (Accessed 3rd April 2019).
- [17] C. Betancur, Etiological heterogeneity in autism spectrum disorders: more than 100 genetic and genomic disorders and still counting, *Brain Res.* 1380 (2011) 42–77.
- [18] K.E. Tansey, et al., Common alleles contribute to schizophrenia in CNV carriers, *Mol. Psychiatry* 21 (2016) 1085–1089.
- [19] E.A. Boyle, Y.I. Li, J.K. Pritchard, Leading Edge Perspective An Expanded View of Complex Traits: From Polygenic to Omnigenic, (2017), <https://doi.org/10.1016/j.cell.2017.05.038>.
- [20] X. Liu, Y.I. Li, J.K. Pritchard, Trans effects on gene expression can drive omnigenic inheritance, *Cell* 177 (2019) 1022–1034.e6.
- [21] E. Fombonne, Epidemiological surveys of autism and other pervasive developmental disorders: An update, *J. Autism Dev. Disord.* 33 (2003) 365–382.
- [22] Turner, T. N. et al. Sex-based analysis of De novo variants in neurodevelopmental disorders. *Am. J. Hum. Genet.* 0, (2019).
- [23] B.S. Abrahams, et al., SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs), *Mol. Autism* 4 (2013) 36.
- [24] J. Veenstra-VanderWeele, S.L. Christian, E.H. Cook Jr., Autism as a paradigmatic complex genetic disorder, *Annu. Rev. Genomics Hum. Genet.* 5 (379–405) (2004).
- [25] E.J. Gardner, et al., Contribution of retrotransposition to developmental disorders, *Nat. Commun.* 10 (2019) 1–10.
- [26] S.J. Sanders, et al., Insights into autism Spectrum disorder genomic architecture and biology from 71 risk loci, *Neuron* 87 (2015) 1215–1233.
- [27] J. Sebat, et al., Large-scale copy number polymorphism in the human genome, *Science* 305 (2004) 525–528.
- [28] A.J. Iafrate, et al., Detection of large-scale variation in the human genome, *Nat. Genet.* 36 (2004) 949–951.
- [29] M. Zarrei, et al., A large data resource of genomic copy number variation across neurodevelopmental disorders, *NPJ Genomic Med.* 4 (2019) 1–13.
- [30] D. Botstein, N. Risch, Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease, *Nat. Genet.* 33 (2003) 228–237.
- [31] R.K. Yuen, et al., Genome-wide characteristics of de novo mutations in autism, *NPJ Genomic Med.* 1 (2016), <https://doi.org/10.1038/npjgenmed.2016.27>.
- [32] J. Zhou, et al., Whole-genome deep-learning analysis identifies contribution of

- noncoding mutations to autism risk, *Nat. Genet.* 51 (2019) 973–980.
- [33] E.K. Ruzzo, et al., Inherited and De novo genetic risk for autism impacts shared networks, *Cell* 178 (2019) 850–866.e26.
- [34] W.M. Brandler, et al., Paternally inherited cis-regulatory structural variants are associated with autism, *Science* 360 (2018) 327–331.
- [35] D.C. Glahn, et al., Rediscovering the value of families for psychiatric genetics research, *Mol. Psychiatry* 24 (2019) 523–535.
- [36] J. Sebat, et al., Strong association of De novo copy number mutations with autism, *Science* 316 (2007) 445–449.
- [37] D. Pinto, et al., Functional impact of global rare copy number variation in autism spectrum disorders, *Nature* 466 (2010) 368–372.
- [38] M. Ronemus, I. Iossifov, D. Levy, M. Wigler, The Role of De Novo Mutations in the Genetics of Autism Spectrum Disorders, (2014), <https://doi.org/10.1038/nrg3585>.
- [39] L. Klei, et al., Common genetic variants, acting additively, are a major source of risk for autism, *Mol. Autism* 3 (2012) 9.
- [40] P. Feliciano, et al., Exome sequencing of 457 autism families recruited online provides evidence for autism risk genes, *NPJ Genomic Med.* 4 (2019) 1–14.
- [41] V.M. Leppa, et al., Rare inherited and De novo CNVs reveal complex contributions to ASD risk in multiplex families, *Am. J. Hum. Genet.* 99 (2016) 540–554.
- [42] R.K.C. Yuen, et al., Whole-genome sequencing of quartet families with autism spectrum disorder, *Nat. Med.* 21 (2015) 185–191.
- [43] M. Woodbury-Smith, et al., Variable phenotype expression in a family segregating microdeletions of the NRXN1 and MBD5 autism spectrum disorder susceptibility genes, *NPJ Genomic Med.* 2 (2017) 1–8.
- [44] C.S. Leblond, et al., Both rare and common genetic variants contribute to autism in the Faroe Islands, *NPJ Genomic Med.* 4 (2019) 1.
- [45] C.P. Schaaf, et al., A framework for an evidence-based gene list relevant to autism spectrum disorder, *Nat. Rev. Genet.* (2020) 1–10, <https://doi.org/10.1038/s41576-020-0231-2>.
- [46] D.B. Callaghan, et al., Whole genome sequencing and variant discovery in the ASPIRE autism spectrum disorder cohort, *Clin. Genet.* (2019), <https://doi.org/10.1111/cge.13556>.
- [47] C.F. Wright, et al., Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data, *Lancet (London, England)* 385 (2015) 1305–1314.
- [48] M.J. Landrum, et al., ClinVar: public archive of relationships among sequence variation and human phenotype, *Nucleic Acids Res.* 42 (2014).
- [49] C.A. Steward, et al., Re-annotation of 191 developmental and epileptic encephalopathy-associated genes unmasks de novo variants in SCN1A, *NPJ Genomic Med.* 4 (2019).
- [50] M.T.W. Ebbert, et al., Systematic analysis of dark and camouflaged genes reveals disease-relevant genes hiding in plain sight, *Genome Biol.* 20 (2019) 1–23.
- [51] J.P. Casey, et al., A novel approach of homozygous haplotype sharing identifies candidate genes in autism spectrum disorder, *Hum. Genet.* 131 (2012) 565–579.
- [52] C. Lord, et al., Autism spectrum disorder, *Nat. Rev. Dis. Prim.* 6 (2020) 1–23.
- [53] M.-C. Lai, et al., Prevalence of co-occurring mental health diagnoses in the autism population: a systematic review and meta-analysis, *Lancet Psychiatry* 6 (2019) 819–829.
- [54] J.Y. An, C. Claudianos, Genetic heterogeneity in autism: from single gene to a pathway perspective, *Neurosci. Biobehav. Rev.* 68 (2016) 442–453.
- [55] L.S. Carroll, M.J. Owen, Genetic overlap between autism, schizophrenia and bipolar disorder, *Genome Med.* 1 (2009) 102.
- [56] D.H. Geschwind, J. Flint, Genetics and genomics of psychiatric disease, *Science* 349 (2015) 1489–1494.
- [57] P.H. Lee, et al., Genomic relationships, novel loci, and pleiotropic mechanisms across eight psychiatric disorders, *Cell* 179 (2019) 1469–1482.e11.
- [58] S.H. Lee, et al., Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs, *Nat. Genet.* 45 (2013) 984–994.
- [59] H.O. Heyne, et al., De novo variants in neurodevelopmental disorders with epilepsy, *Nat. Genet.* 50 (2018) 1048–1053.
- [60] K.S. Barton, et al., Pathways from autism spectrum disorder diagnosis to genetic testing, *Genet. Med.* 20 (2018) 737–744.
- [61] S.M. Myers, et al., Insufficient evidence for “autism-specific” genes, *Am. J. Hum. Genet.* 106 (2020) 587–595.
- [62] S. Srivastava, et al., Meta-analysis and multidisciplinary consensus statement: exome sequencing is a first-tier clinical diagnostic test for individuals with neurodevelopmental disorders, *Genet. Med.* 1 (2019), <https://doi.org/10.1038/s41436-019-0554-6>.
- [63] K. Tammimies, et al., Molecular diagnostic yield of chromosomal microarray analysis and whole-exome sequencing in children with autism Spectrum disorder, *JAMA* 314 (2015) 895.
- [64] C. Gillissen, et al., Genome sequencing identifies major causes of severe intellectual disability, *Nature* 511 (2014) 344–347.
- [65] T.N. Turner, et al., Genome sequencing of autism-affected families reveals disruption of putative noncoding regulatory DNA, *Am. J. Hum. Genet.* 98 (2016) 58–74.
- [66] T.N. Turner, et al., Genomic patterns of De novo mutation in simplex autism, *Cell* 171 (2017) 710–722.e12.
- [67] S. Lee, J.G. Gleeson, Closing in on mechanisms of open neural tube defects, *Trends Neurosci.* 43 (2020) 519–532.
- [68] L. D'Abate, et al., Predictive impact of rare genomic copy number variations in siblings of individuals with autism spectrum disorders, *Nat. Commun.* 10 (2019) 5519.
- [69] I. Mitra, et al., The contribution of de novo tandem repeat mutations to autism spectrum disorders, *bioRxiv* (2020), <https://doi.org/10.1101/2020.03.04.974170>.
- [70] N. Mousavi, S. Shleizer-Burko, R. Yanicky, M. Gymrek, Profiling the genome-wide landscape of tandem repeat expansions, *Nucleic Acids Res.* 47 (2019) 90.
- [71] J.Y. An, et al., Towards a molecular characterization of autism spectrum disorders: An exome sequencing and systems approach, *Transl. Psychiatry* 4 (2014).
- [72] J.D. Buxbaum, et al., The autism sequencing consortium: large-scale, high-throughput sequencing in autism spectrum disorders, *Neuron* 76 (2012) 1052–1056.
- [74] C.B. Pedersen, et al., The iPSYCH2012 case-cohort sample: new directions for unravelling genetic and environmental architectures of severe mental disorders, *Mol. Psychiatry* 23 (2018) 6–14.
- [75] F.K. Satterstrom, et al., ASD and ADHD have a similar burden of rare protein-truncating variants, *bioRxiv* (2018), <https://doi.org/10.1101/277707>.
- [76] R.K.C. Yuen, et al., Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder, *Nat. Neurosci.* 20 (2017) 602–611.
- [77] D. Levy, et al., Rare De novo and transmitted copy-number variation in autistic Spectrum disorders, *Neuron* 70 (2011) 886–897.
- [78] S.J. Sanders, et al., Multiple recurrent De novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism, *Neuron* 70 (2011) 863–885.
- [79] J.Y. An, et al., Genome-wide de novo risk score implicates promoter variation in autism spectrum disorder, *Science* 362 (2018).
- [80] D.M. Werling, et al., An analytical framework for whole-genome sequence association studies and its implications for autism spectrum disorder, *Nat. Genet.* 50 (2018) 727–736.
- [81] T.A.G.P. Consortium, et al., Mapping autism risk loci using genetic linkage and chromosomal rearrangements, *Nat. Genet.* 39 (2007) 319–328.
- [82] J.D. Buxbaum, et al., The autism simplex collection: an international, expertly phenotyped autism sample for genetic and phenotypic analyses, *Mol. Autism* 5 (2014) 34.
- [83] A.A. Eshraghi, et al., Epigenetics and autism spectrum disorder: is there a correlation? *Front. Cell. Neurosci.* 12 (2018) 78.
- [84] A.L. McGuire, et al., The road ahead in genetics and genomics, *Nat. Rev. Genet.* (2020) 1–16, <https://doi.org/10.1038/s41576-020-0272-6>.