1. Introduction

Although autism is considered to be one of the most highly heritable psychiatric disorders, molecular mechanisms underlying its pathogenesis remain largely unresolved. A strong genetic component underlying autism spectrum disorders (ASD) has been firmly established from various lines of studies ranging from whole genome scans to genetic association studies. Recent genomic advances have led to steep growth in the number of diverse genetic loci linked to ASD, including candidate genes containing rare or common variants, chromosomal aberrations, and submicroscopic copy number variations. Additionally, autism is consistently associated with a number of single gene mutation disorders such as Fragile X Syndrome. Most genetic variations fail to replicate between studies and populations, further complicating our understanding of ASD disease etiology. Here we review recent expansion of heterogeneity in the genetic landscape for ASD. First we define the types of genetic risk factors implicated in this disorder. We then comparatively analyze the pools of ASD candidate genes identified as of the end of years 2006 and 2010, profiling both their distribution and molecular function. We highlight bioinformatics tools for ASD which can be used to build and evaluate networks of ASD genes as the number of risk factors grows. Finally, we discuss the impact of genetic heterogeneity on theories of ASD pathogenesis.

2. Genes

In the post-genomic era, continuous identification of new ASD risk factors has rapidly expanded the types of candidate genes implicated in the pathogenesis of this disorder. Until 2003, single gene mutations in ASD were derived from well-characterized genetic syndromes such as Fragile X Syndrome and Rett Syndrome, in which subpopulations of individuals develop autistic symptoms. Later that year, Thomas Bourgeron’s group first identified single gene mutations/disruptions in neuroligins in siblings with ASD (Jamain et al., 2003). This seminal work opened up the field of ASD research in two major areas: first, a strong genetic foundation to non-syndromic forms of ASD and, second, a focus on the synaptic model for the disorder. Since then, high throughput genetic studies have rapidly identified additional genetic risk factors, vastly expanding the pool of ASD-linked genes.
Candidate genes for ASD can currently be defined into four distinct sets:

1. **Rare**: genes implicated in rare monogenic forms of ASD. The types of allelic variants within this class include rare polymorphisms and single gene disruptions/mutations directly linked to ASD. Examples include NRXN1 and SHANK3.

2. **Syndromic**: genes implicated in syndromes in which a significant subpopulation develops autistic symptoms. Examples include FMR1 (Fragile X Syndrome) and MECP2 (Rett Syndrome).

3. **Association**: genes with common polymorphisms that confer small risk for ASD and have been identified from genetic association studies of ASD derived from unknown cause (known as “idiopathic ASD”). Examples include MET and GABRB1.

4. **Functional**: genes with functions relevant for ASD biology and not included in any of the other genetic categories. Examples include CADSP2, for which knockout mouse models exhibit autistic characteristics, but the gene itself has not been directly tied to known cases of ASD.

Of these four gene categories, *Rare* and *Syndromic* contain the strongest evidence of links to ASD (for review, El-Fishawy & State, 2010). *Association* genes lack replication of their relationship to ASD, and *Functional* genes have no documented direct link to ASD. Over 200 ASD candidate genes have been reported thus far in the scientific literature (Table 1). These genes are distributed at discrete regions throughout the entire genome (Figure 1).

Fig. 1. Ideogram of all currently known ASD candidate genes. To the left of each chromosome are genes that fall within either the *association* category (red) or the *functional* category (black), whereas to the right of each chromosome are genes that fall within either the *rare* mutation category (green) or the *syndromic* category (blue). * = gene also identified by genome-wide association studies.

### 3. Chromosomes

Microscopically visible large-scale chromosomal rearrangements have long been implicated in the onset and progression of a host of developmental disorders. Deletions of the 15q11-q13 region on the maternal chromosome lead to Angelman syndrome (Williams et al., 2008), whereas the corresponding deletion on the paternal chromosome gives rise to Prader-Willi syndrome (Cassidy & Schwartz, 2009). Deletions, duplications, translocations, and
inversions larger than 3 Mb responsible for these and other syndromes have traditionally been identified by microscopic techniques such as karyotyping and, more recently, fluorescent in situ hybridization (FISH). In recent years, technological and computational advances have provided researchers with the sensitivity and accuracy to identify structural variation in chromosomes less than 3 Mb in size, which could not have previously been identified by traditional cytogenetic methods such as karyotyping.

<table>
<thead>
<tr>
<th>Genetic Category</th>
<th>Number of Genes</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rare</td>
<td>81</td>
<td>ANKRD11, A2BP1*, APC*, ASTN2, AUTS2, BZRAP1, C3orf58, CA6, CACNA1H, CADM1, CENTG2*, CNTN4, CNTNAP2*, CNTNAP5, CXXR3, DIAPH3, DLGAP2, DPP10, DPP6, DPD6, EIF4E, FABP5*, FABP7*, FBXO40, FHT1, FRMPD4, GALNT13, GLRA2, GRPR, HNRNPH2, IL1RAPL1, IMPM2L*, JMJD1C, KCNMA1, KIAA1586, MBD1, MBD3, MBD4, MCPH1, MDGA2, MEF2C, NBEA, NLGN1, NLGN3, NLGN4X, NOS1AP, NRXN1, ODF3L2, OPHN1, ORC1, PARK2, PCDH9, PCDH10, PCDH19, PDZD4, PLN, PPP1R3F, PSMD10, PTCHD1, RAB39B*, RAPGEF4, RB1CC1, REEP3, RFWD2, RIMS3, RPL10, RS6KA2, SCN1A, SCN2A, SEZ6L2, SH3BP1, SHANK2, SHANK3, SLC4A10, SLC9A9, ST7, SUCLG2, TMEM195, TSPAN7, UBE3A*, WNK3</td>
</tr>
<tr>
<td>Syndromic</td>
<td>21</td>
<td>ADSL, AGTR2, AHI1*, ALDH5A1, ARX, CACNA1C, CACNA1F, CDKL5, DHCRC7, DMD, DMPK, FMR1, MECP2, NF1, NTNG1, PTEN, SLC6A8, SLC9A6, TSC1, TSC2, XPC</td>
</tr>
<tr>
<td>Association</td>
<td>84</td>
<td>ABAT, ADA, ADORA2A, ADRB2, AR, ARNT2, ASMT, ATP10A, AVPR1A, C4B, CACNA1G, CCDC64, CDH10, CDH22, CDH9, CTNNAS, CYP11B1, DISC1, DLX1, DLX2, DRD3, EN2, ESR1, ESRRB, FBXO33, FEZ1F2, FOXP2, FRK, GABRA4, GABRB1, GABRB3, GLO1, GPX1, GRK2, GRIN2A, GRM8, GSTM1, HLA-A, HLA-DRB1, HOXA1, HRAS, HS3ST5, HSD11B1, HTR1B, HTR3A, HTR3C, INPP1, ITGA4, ITGB3, LAMB1, LRFN5, LRRC1, LZTS2, MACROD2, MARK1, MET, MTF1, MYO16, NOS2A, NPA52, NRCAM, NR2P2, NTRK1, NTRK3, OXTR, PER1, PIK3CG, PITX1, PON1, PRKCB1, PTGS2, RELN, RHOXF1, SLC1A1, SLC25A12, SLC6A4, STK39, SYT17, TDO2, TPH2, UBE2H, VASH1, WNT2</td>
</tr>
<tr>
<td>Functional</td>
<td>23</td>
<td>ALOX5AP, ASS, CACNA1D, CADPS2, CBS, CD44, CNR1, DAB1, DAPK1, DCUN1D1, DDX11, EGR2, F13A1, FLT1, ITGB7, MAOA, MAP2, OPRM1, RA1I, ROBO1, SDC2, SEMA5A, TSN</td>
</tr>
</tbody>
</table>

Table 1. Genetic classification of all currently identified ASD candidate genes. (* = gene replicated by independent association studies.)
Of particular interest in the field of submicroscopic structural variants are deletions and duplications collectively categorized as copy number variants. A copy number variant (CNV) is typically defined as a ≥1 kb DNA segment that is present at a differing copy number compared to a reference genome (Feuk et al., 2006). CNVs can either arise de novo or be inherited on the maternal and/or paternal chromosome. Much like many single nucleotide polymorphisms, apparently benign CNVs exist in the general population at relatively high frequencies; as such, CNVs that exist in the general population at a rate of 1% or higher are generally described as CNV polymorphisms (Feuk et al, 2006). Submicroscopic copy number variants have come under increased scrutiny in recent years as a potential causative agent in the onset and progression of developmental disorders, including neuropsychiatric disorders such as ASD.

3.1 Copy number variation in autism spectrum disorders
As more syndromes were subsequently shown to be associated with both microscopic and submicroscopic chromosomal structural variation, it became apparent that a subset of patients diagnosed with some of these syndromes also developed ASD or displayed autistic traits. For example, DiGeorge Syndrome (also called Velocardiofacial Syndrome), which is frequently characterized by congenital heart anomalies, palatal abnormalities, immune system deficits and some degree of facial dysmorphism, has been found to result from a ~3 Mb deletion in chromosome 22 (McDonald-McGinn et al., 2005). Individuals diagnosed with this syndrome, also referred to as 22q11.2 deletion syndrome, frequently experience learning disabilities; however, approximately 20% of patients with this syndrome also develop ASD. Given that a subset of patients with syndromes caused by chromosomal structural abnormalities also display autistic traits, as well as the high prevalence of ASD in individuals with cytogenetically visible duplications of the Angelman/Prader-Willi syndrome region (15q11-q13) on the maternal chromosome (Cook Jr. et al., 1997; Schroer et al., 1998), a number of studies in the past decade have focused on identifying submicroscopic structural variants, in particular CNVs, in individuals with ASD and subsequently determining the importance of these variants in disease pathogenesis. In order to more fully ascertain the pathogenic risk associated with copy number variants, only patients with idiopathic cases of ASD have typically been used; patients with mutations in genes previously implicated in ASD, such as the \textit{FMR1} gene, or with gross chromosomal abnormalities have frequently been excluded from these studies.

The advent of genome-wide scanning technologies has enabled researchers to identify and subsequently confirm >1200 potentially pathologically relevant CNVs located within over 490 distinct loci in autistic populations since 2007 (Sebat et al., 2007; Szatmari et al., 2007; Marshall et al., 2008; Cuscó et al., 2009; Glessner et al., 2009; Gregory et al., 2009; van der Zwaag et al., 2009; Pinto et al., 2010; Bremer et al., 2011). Confirmation or validation of a CNV by an independent approach following its discovery is essential not only to remove false positives, but also to more accurately identify the boundaries of a CNV. Validated CNVs in autistic individuals have been located in loci on all 22 somatic chromosomes and the X chromosome (Figure 2).

While many of the CNVs identified by these methods are singletons and require additional replication to more accurately assess their potential role in disease, there are rare, recurring CNVs at particular loci that have been identified across multiple autistic populations that have emerged as strong risk-conferring candidates in ASD pathogenesis. Ten loci that have been identified multiple times in autistic case populations are described in Table 2. Perhaps the most intensely studied of these recurring CNVs, aside from duplications in the 15q11-13...
loci, are ~500 kb deletions and duplications that occur at the 16p11.2 locus. A recently published meta-analysis of the 16p11.2 locus in autistic populations discovered that CNVs at the 16p11.2 locus have a prevalence of 0.76%, with deletions occurring approximately twice as frequently with duplications (Walsh & Bracken, 2011). CNVs in autistic individuals have been identified in regions previously associated with other deletion-duplication syndromes, such as the 1q21.1, 22q11.21 and 22q13.33 loci (McDonald-McGinn et al., 2005; Phelan, 2007; Haldeman-Englert & Jewett, 2011). Other strong candidate CNV loci to emerge from genome-wide scanning assays include 2p16.3, 3p26.3, 6q26, 7q11.22, and 15q13.3. In some cases, CNVs at these “hot-spot” loci appear to target genes that have previously been implicated in ASD pathogenesis, such as NRXN1 (2p16.3), PARK2 (6q26), and AUTS2 (7q11.22).

![Fig. 2. Validated copy number variants (CNVs) identified in genome-wide scanning arrays from nine published reports (Sebat et al., 2007; Szatmari et al., 2007; Marhall et al., 2008; Cuscó et al., 2009; Glessner et al., 2009; Gregory et al., 2009; van der Zwaag et al., 2009; Pinto et al., 2010; Bremer et al., 2011). The red lines to the left of each chromosome represent the >1200 validated CNVs identified in these studies. Especially long CNVs that overlap smaller CNVs are represented with thinner red lines. The green lines to the right of chromosomes 1, 2, 3, 6, 7, 15, 15, and 22 represent ten rare, recurring CNV loci identified in at least three of the nine aforementioned publications (described in more detail in Table 2).

Increasingly, targeted assays using methods such as quantitative PCR are being used to characterize CNVs at particular loci that have been previously identified by more global scanning approaches, given the relatively high frequency of these CNVs in autistic case cohorts. CNVs are now considered one of the most common, genetic causes of ASD, with 10-20% of ASD cases believed to be the result of submicroscopic deletions and duplications (Miles et al., 2010).
Table 2. Examples of loci in which rare, recurring, potentially risk-conferring CNVs are frequently observed in autistic populations. Potential candidate genes within the locus, the type of CNV that targets each locus, and the articles in which these CNV loci were identified are included. *, CNV locus which overlaps with adjacent loci in at least one of the publications listed.

A more detailed analysis of the nine published research articles used to construct the ideogram in Figure 2 reveals that, while the percentage of previously unidentified CNV loci has steadily declined since 2007, new CNV loci still constitute a very high percentage of the total CNV loci identified and validated in these studies (Table 3). Therefore, while recurring CNVs such as 16p11.2 and others continue to be observed across multiple autistic populations and CNV studies, novel CNVs in autistic populations are still being identified, indicating that there are likely multiple potential targets for the pathogenic properties of CNVs throughout the human genome. It is likely that other novel CNVs in autistic individuals have not yet been identified, and as such their identification will shed new light on the pathways adversely affected in ASD.

Table 3. Analysis of the nine papers used to construct the ideogram in Figure 2. While the overall percentage of previously unidentified CNV loci has decreased from year to year, novel CNV loci still constitute the majority of the total CNV loci that have been identified and confirmed in these studies.
3.2 Risk-conferring vs. benign copy number variants
Although advances in genome-wide and targeted scanning assays have enabled researchers to discover potentially risk-conferring CNVs in autistic individuals, significant issues remain in the determination of which CNVs are pathologically relevant or benign in nature. This is of particular importance in terms of potentially using genetic screening for risk-conferring CNVs as a tool to assess the risk of ASD in unborn children. The diagnostic accuracy of such a screening protocol would be entirely dependent on knowing which CNVs would confer the greatest potential risks for ASD pathogenesis. In order to distinguish between risk-conferring and benign CNVs in an autistic population, a comparison must be made between both the existence and frequency of CNVs between affected and unaffected individuals. To account for possible genetic differences between ethnic groups, it is critical that a control population of comparable size and ethnic background be included in any CNV study. For example, CNVs at loci thought to confer a high risk of ASD susceptibility, such as deletions and duplications at the 16p11.2 locus, have also been identified in healthy individuals, although at a much lower frequency than in autistic populations. Given the increased frequency of CNVs at the 16p11.2 loci in autistic populations versus control populations, CNVs at this region remain classified as high risk-conferring CNVs. In addition, there are online databases such as the Database of Genomic Variants (http://projects.tcag.ca/variation/) and the Copy Number Variant resource at the Children’s Hospital of Philadelphia (http://cnv.chop.edu/) available that describe previously identified CNVs in healthy individuals. These tools provide a means to further filter out likely benign CNVs from autistic case studies and enrich for potentially pathogenic variants. However, it should be noted that seemingly benign CNVs may be involved in more subtle phenotypes in autistic individuals when occurring in combination with other factors. Likewise, additional meta-analysis studies of CNV loci across multiple published autistic populations, such as that described for the 16p11.2 locus, will be required to compare frequencies of CNV a in order to more fully determine the global risk potential associated with any given CNV at a particular locus.

3.3 De novo vs. inherited copy number variants
As previously stated, CNVs can either arise de novo, or be inherited from the mother and/or father. Considerable interest has been placed in the pathogenic importance of de novo CNVs as a cause of ASD compared to inherited variants, especially within the context of sporadic vs. familial ASD cases. Indeed, some studies have found that the rate of de novo CNVs is higher in sporadic cases compared to familial cases (Sebat et al., 2007; Marshall et al., 2008), while Bremer et al. (2011) found that the rate of rare inherited CNVs was higher in familial cases compared to sporadic cases. These findings would suggest that de novo CNVs are predominantly responsible for ASD in sporadic cases, whereas inherited CNVs are primarily responsible for familial cases of ASD. However, Pinto et al. (2010) found no significant difference between the frequencies of de novo CNVs in sporadic vs. familial cases. It has been reported that validated de novo CNVs strongly associate with ASD (Sebat et al., 2007). However, there is no firm evidence that de novo CNVs confer a higher probability or severity of disease than inherited variants. On the other hand, the dynamics of CNV inheritance and subsequent susceptibility to ASD has its own issues: an autistic individual with a potential risk-conferring CNV may inherit that CNV from a parent who fails to exhibit autistic traits; an autistic individual may have unaffected siblings who have likewise inherited the identical CNV; or one affected sibling in a multiple family may have a risk-conferring CNV, whereas other affected siblings may not.
3.4 Copy number variation and phenotypic heterogeneity

Detailed studies attempting to correlate genotype with phenotype have demonstrated that there is significant phenotypic heterogeneity between individuals with CNVs at a particular chromosomal locus, both in terms of disease presence and severity of disease. Studies in autistic populations containing CNVs at the 15q13.3 (Miller et al., 2009; Ben-Shachar et al., 2009) and 16p11.2 (Fernandez et al., 2010) loci, for example, have shown that autistic phenotypes, such as the extent of facial dysmorphism and the extent of intellectual disability, can vary from one patient to the next with the same CNV. One model that has been designed to address some of the issues as to how CNVs contribute to ASD states that certain CNVs at particular loci increase the susceptibility of an individual to developing an ASD based on a “threshold” of disease severity (Cook & Scherer, 2008). Chief among these high susceptibility CNVs are maternal duplications at 15q11-q13, deletions at 16p11.2, and deletions at the loci encoding for cell adhesion proteins such as neuroligins. Other rare recurring CNVs that have been identified in autistic populations may confer a lower overall risk of ASD pathogenesis, or a decreased severity of disease, such as CNVs at 1q21.1, 2p16.3, and 22q11.21. However, even these CNVs can result in the onset of ASD, or more severe disease phenotypes, when in combination with other genetic and non-genetic factors. These genetic factors may include additional CNVs (indeed, many autistic individuals have more than one CNV within their genome) or single gene mutations, such as those described elsewhere in this chapter, whereas non-genetic factors can be environmental, sex-related, or epigenetic in nature. Epigenetic regulation of gene expression may be of particular importance with regards to phenotypic heterogeneity in autistic individuals with 15q11-q13 duplications, as this region contains a number of potentially critical imprinted genes. Further studies involving more detailed analysis of genotype-phenotype correlations in autistic individuals with CNVs will be instrumental in determining the role of CNVs in ASD.

3.5 Mechanism of action of copy number variants

The general mechanism by which a CNV might contribute to ASD pathogenesis remains unclear. The simplest mechanism of action involves gene dosage, by which deletion or duplication of a gene or genes within a particular CNV locus, or the deletion or duplication of gene regulatory elements, subsequently results in altered or disrupted levels of gene product. A deletion at a particular locus might also result in the unmasking of a recessive gene on the corresponding chromosomal locus, which would then be able to elicit a deleterious effect. Such a mechanism might be involved in disease pathogenesis in an autistic individual with a 10 Mb maternally inherited deletion in chromosome 13q and a point mutation in the DIAPH3 gene on the paternal chromosome (Vorstman et al., 2010). As the proband’s unaffected sibling also had the DIAPH3 mutation, but lacked the corresponding deletion, it is tempting to argue that the maternal deletion unmasked a recessive mutation in the paternal DIAPH3 gene, and that in turn influenced the onset of ASD in the proband. Given that many CNVs are large enough to include up to 50 or more genes, identifying which genes are of functional relevance in ASD pathogenesis within a particular CNV loci remains a challenging task. Much in the same way that genes that confer susceptibility to ASD have been found to fall within intriguing functional categories, bioinformatic analysis of genes that lie within or adjacent to recurring CNV loci may yield similar results and aid in both identifying new candidate genes and in discovering conserved pathways potentially targeted by copy number variation. Indeed, analysis designed to identify potentially relevant functional pathways containing genes located in copy number variants have been performed (Pinto et al., 2010).
4. Comparative analysis of ASD genes

To analyze recent evolution of the ASD molecular landscape, we profiled ASD genes identified as of the end of years 2006 and 2010. To define pools of ASD candidate genes existing at these time points, we used the ASD database AutDB (www.mindspec.org/autdb.html), a publicly available, curated, web-based, searchable genetic database for ASD created by our laboratory (see Section 5). We then examined the genetic and functional expansion of these gene sets.

4.1 Genetic expansion

To quantify the total number of ASD candidate genes identified as of 2006 and 2010, we sorted existing ASD candidate genes according to year of first publication. We discovered that the total number of ASD candidate genes more than doubled in the past four years: whereas 91 genes were linked to ASD as of 2006, this number rapidly grew to 209 genes in 2010.

To compare genetic distribution within these datasets, we defined ASD candidate genes according to the classification system described in Section 2: rare variants (Rare), syndromic genes (Syndromic), genes identified by association studies (Association), and genes whose functions have been implicated in ASD (Functional). We found that expansion of the total ASD gene pool was largely due to steep growth of both Rare and Association gene sets, with a slight increase in the numbers of identified Syndromic and Functional genes (Figure 3). Notably, the near quadrupling of the number of rare mutations supports the Rare Allele, Common Disease as a plausible theory of ASD pathogenesis (see Section 6).

![Fig. 3. Genetic distribution of ASD candidate genes identified as of the end of years 2006 and 2010.](image-url)

4.2 Functional expansion

Recent large-scale ASD studies have used a systems biology approach to translate genetic information into functional maps. For instance, Glessner et al. (2009) showed that ASD-linked genes cluster in synaptic processes such as cell adhesion and ubiquitin-mediated
degradation. Additionally, in the largest ASD study performed to date, Pinto et al. (2010) found that genes affected by rare CNVs were enriched in functions such as neuronal development and GTPase/Ras signaling.

To build upon these functional maps, we used a well-known synaptic proteome classification system (Husi et al., 2000), to organize ASD gene sets from 2006 and 2010 into eight broad categories of molecular function, defined by corresponding subcategories:

1. **Cell Adhesion** (cell adhesion molecule, cell adhesion/axon guidance, extracellular matrix, extracellular secreted protein)
2. **Guidance/Outgrowth** (axon guidance, cell migration, cell surface glycoprotein, cytoskeletal remodeling, dendritic spine morphology, animal model evidence)
3. **Neurotransmission** (adaptor protein, G-protein coupled receptors, ligand-gated ion channel, neuromodulator receptor, neuromodulator receptor-associated protein, neuromodulator synthesis, neurotransmitter receptor, neurotransmitter synthesis, presynaptic release, scaffolding protein, sensory receptor, transporter, voltage-gated ion channel, voltage-gated ion channel modulator)
4. **Signaling** (glycosylation, kinase, kinase substrate, phosphatase, proteoglycan, small G-protein or modulator, tyrosine receptor kinase, other signal)
5. **Degradation** (proteasome-related protein, ubiquitin ligase)
6. **Transcription** (circadian protein, cofactor, DNA binding, DNA damage response protein, DNA methylation, estrogen receptor, histone demethylation protein, homeodomain protein, preinitiation complex, purine metabolism, transcription factor)
7. **Translation** (ribosomal protein, RNA binding, RNA metabolism, RNA structure)
8. **Other** (antioxidant, endosome regulation, energy production, fatty acid binding protein, immune system, membrane biosynthesis, mitochondrial carrier protein, mitochondrial targeting protein, oxidation, prostaglandin, unknown function).

The functional distribution of ASD risk genes vastly expanded from 2006 to 2010 (Figure 4). Because Rare and Syndromic genes contain the strongest links to ASD (see Section 2), we examined this combined “Rare/Syndromic” set as one dataset. We comparatively assessed them with Association genes as a separate gene set. Both Rare/Syndromic and Association gene datasets followed the same trend: whereas Neurotransmission and Signaling were by far the largest functional categories in 2006, the number of genes in all other functional categories increased over the past four years such that all are becoming relatively equalized. The most dramatic increases occurred in Cell Adhesion, Degradation, Transcription, and Other.

This functional expansion has led to shifting theories of ASD pathogenesis. In 2006, the largest percentage of ASD susceptibility genes resided in the Neurotransmission or Signaling categories, supporting specific theories of dysfunction, such as serotonin transport (Cook & Leventhal, 1996). However, rapid expansion of nearly all functional categories throughout 2010 indicates that ASD susceptibility genes are actually widespread in neurobiological function. Such functional expansion supports broad theories of pathogenesis such as the proposed enhancement of brain excitability in ASD (Rubenstein & Merzenich, 2003). Each designated functional category includes neurobiological factors that contribute to brain excitability, reinforcing the idea that mutations in vastly different genes may facilitate similar outcomes in brain function by contributing to shared molecular pathways. Together, accelerated identification of ASD risk genes with widespread neurobiological functions is leading to a convergent model of ASD pathogenesis.
5. Bioinformatics of ASD

The enormous amount of data currently being generated by large-scale genomic studies poses a critical challenge for its storage and analysis. To process this information, bioinformatics tools are becoming increasingly vital to the scientific community. Here we highlight several ASD-related databases which researchers can use to navigate this data and shed insight into the molecular pathways underlying ASD pathogenesis.

5.1 AutDB

Our laboratory created the ASD database AutDB (www.mindspec.org/autdb.html), the first publicly available, curated, web-based, searchable genetic database for ASD (Basu et al., 2009; Kumar et al., 2011). In AutDB, evidence regarding ASD candidate genes is systematically extracted from peer-reviewed, primary scientific literature and manually curated by our researchers. To provide high-resolution view of various components linked to ASD, we developed detailed annotation rules based on the biology of each data type and generated controlled vocabulary for data representation. AutDB is widely used by individual laboratories (Crespi et al., 2010; Elia et al., 2010; Gillis et al., 2010; Toro et al., 2010) and consortiums (Simons Foundation) for understanding genetic bases of ASD.

With a systems biology approach, AutDB integrates various modules encompassing different types of data relevant for ASD:

**Human Gene:** This original module of AutDB includes all genes whose mutations have been associated or implicated with ASD, together with all risk-conferring candidates.
associated with these disorders (Basu et al., 2009). ASD-related genes are classified into the four categories described in Section 2: 1) Rare: genes implicated in rare monogenic forms of ASD; 2) Syndromic: genes implicated in syndromic forms of ASD where a subpopulation with a specific genetic syndrome develops autistic symptoms; 3) Association: small risk-conferring candidate genes with common polymorphisms identified from genetic association studies in idiopathic ASD; and, 4) Functional: candidates genes with functions relevant for ASD biology, not covered by any of the previous genetic categories. All known ASD-specific mutations at the DNA sequence level will be available by late 2011.

Animal Model: This module provides a comprehensive collection of all mouse models linked to ASD (Kumar et al., 2011). The core behavioral features of ASD involve higher order human brain functions like social interactions and communications, which can only be approximated in animal models, so the annotation strategy for this module includes four broad areas: 1) core behavioral features of ASD, 2) ASD-related traits such as seizures and circadian rhythms that are heritable and more easily quantified in animal models; 3) neuroanatomical features, and 4) molecular profiles. To this end, we developed PhenoBase, a classification table for systematically annotating models with controlled vocabulary. PhenoBase contains 16 major categories and >100 standardized phenotype terms.

Protein Interaction (PI): This module serves as a repository for all known protein-protein interactions of ASD candidate genes. It documents five major types of direct interactions: 1) protein binding, 2) promoter binding, 3) RNA binding, 4) protein modification, and 5) direct regulation. One of the newest additions to AutDB, a beta version of this module was released in April 2011, with a full version scheduled for release in late 2011. Its content is envisioned to have immediate application for network biology analysis of molecular pathways involved in ASD pathogenesis.

Copy Number Variant (CNV): This module is a comprehensive, up-to-date reference for all known copy number variants (CNVs) implicated in ASD (see Section 3). It originates from a multi-level annotation model including data such as chromosomal location, size, and relevance to ASD. Like the PPI module, a beta version of the CNV module was released in May 2011, with a full version scheduled for release in late 2011.

5.2 ASD Chromosome Rearrangement Database
The ASD Chromosome Rearrangement Database (http://projects.tcag.ca/ASD/) is a web-based, searchable genetic database hosted by The Centre for Applied Genomics at the Hospital for Sick Children in Toronto, Canada (Marshall et al., 2008). The ASD Chromosome Rearrangement Database provides information not only on submicroscopic CNVs, that have been identified by microarray studies, but also data on microscopic structural variants identified by cytogenetic studies. This database is updated both from published research articles and in-house experimental results.

5.3 ASD Genetic Database
The ASD Genetic Database (http://wren.bcf.ku.edu/) is another web-based, searchable genetic database developed by researchers at the University of Kansas (Matuszek & Talabizadeh, 2009). Much like AutDB and the ASD Chromosome Rearrangement Database, the ASD Genetic Database provides information on genes and CNVs believed to impart susceptibility to ASD. However, this database also includes information on known non-
coding RNAs and chemically-induced fragile sites in the human genome. Non-coding RNAs, such as microRNAs, have come under increased scrutiny with regards to their potential pathogenic role in ASD. For example, the 15q11-q13 region contains a number of small nucleolar RNAs (snoRNAs). Duplication of a region of mouse chromosome 7 that has conserved linkage with human chromosome 15q11-q13 in mouse model of ASD resulted in overexpression of the snoRNA MBI52 (the mouse ortholog of the human snoRNA HBII52), which could potentially alter serotonergic signaling and contribute in part to the autistic traits exhibited by these mice (Nakatani et al., 2009). Spontaneous breakage during DNA replication at rare chromosomal fragile sites may also play a role in the pathogenesis of neuropsychiatric disorders such as ASD. The chromosomal fragile site FRAXA has been implicated in fragile X syndrome, and other fragile sites have been identified that associate with ASD, such as FRA2B, FRA6A, and FRA13A (Smith et al., 2010).

6. Discussion

6.1 Rare vs. common alleles
At the beginning of this decade, few single mutations for ASD had been identified. As of 2003, single mutations in only two genes were known: neuroligins 3 and 4, published in a single report (Jamain et al., 2003). This led to predominance of the Common Allele Common Disease theory, which proposes that ASD is caused by combined effects of multiple common polymorphisms.

However, evidence from two recent major studies led to the emergence of an alternative Rare Allele Common Disease theory for ASD pathogenesis. First, comparative genomic hybridization with subsequent confirmation showed a strong association between de novo CNVs mutations and ASD (Sebat et al., 2007). Second, homozygosity mapping identified numerous single gene mutations in families with ASD (Morrow et al., 2008). According to the Rare Allele Common Disease theory, the genetics underlying complex neuropsychiatric disorders such as ASD is highly heterogeneous. It proposes that ASD is caused by numerous rare, highly penetrant mutations that may even by caused by “private mutations” specific to individual families; a similar theory has been proposed to explain the genetic complexity of schizophrenia (McClellan et al., 2007). The identification of rare variants has more than quadrupled in the past four years (see Section 4), lending credibility to this theory.

At present, it appears that the Rare Allele Common Disease theory is a highly relevant genetic paradigm for ASD and other complex disorders. A few recent papers have identified common variants associated with ASD (Campbell et al., 2006; Wang et al., 2009; Weiss et al., 2009; Anney et al., 2010), but these mutations are still far outnumbered by known rare single gene mutations. With increased availability of various types of sequencing technologies, it is projected that additional rare mutations/variations will be discovered or validated rapidly in upcoming years, making clinical genomics of ASD an option for affected families.

6.2 Prioritization of genetic ASD risk factors
In future, ASD risk genes should be prioritized based on careful definitions at both genetic and functional levels. High priority genes should show evidence for replication or participate in a molecular pathway exhibiting multiple ASD-linked mutations. Examples include the cell adhesion molecule CNTNAP2, a neurexin family member in which both
common and rare variants have been associated with ASD (Arking et al., 2008). CNTNAP2 is also regulated by FOXP2, a candidate ASD gene highly relevant for human language development (Gong et al., 2004). Additionally, the role of synaptic scaffolding proteins in ASD has been strengthened by recent identification of recurrent mutations in SHANK2 (Berkel et al., 2010; Pinto et al., 2010). Furthermore, if multiple genes contribute to syndromic ASD, each gene should only be considered high priority when accompanied by a documented direct genetic link to ASD.

CNVs should likewise be prioritized based on a number of factors. A high risk-conferring structural variant should not only display a high prevalence in autistic populations, but also an enrichment in autistic populations compared to control populations. Meta-analysis studies, such as that previously described for the 16p11.2 locus in multiple autistic case studies (Walsh & Bracken, 2011), will aid greatly in determining which CNVs meet these criteria. Furthermore, a high priority CNV locus should either contain a gene or genes, or the regulatory elements for a gene or genes, which demonstrate potential participation in a molecular pathway exhibiting multiple ASD-linked single gene mutations. CNV loci containing genes that have already been associated with increased risk of autism, such as 2p16.3 (NRXN1), are of particular interest in this regard.

6.3 Synaptic theory of ASD
A hypothesis for ASD as a synaptic disorder is well recognized, largely based on strong evidence from rare mutations in neuroligins, neurexins and SHANK3 (Bourgeron, 2009). Rapid expansion of the ASD risk gene pool has supported this synaptic theory of ASD by identifying rare mutations in numerous additional synapse-related genes, including SHANK2 (Berkel et al., 2010; Pinto et al., 2010) and PTCHD1 (Marshall et al., 2008; Noor et al., 2010; Pinto et al., 2010). Additionally, functional maps generated from large-scale studies of ASD have enriched this synaptic hypothesis of ASD, identifying categories ranging from cell adhesion and ubiquitin-mediated degradation (Glessner et al., 2009) to neuronal development and GTPase/Ras signaling (Pinto et al., 2010).

Our functional profile of all ASD candidate genes identified as of 2010 supports this synaptic hypothesis (see Section 4). The majority of ASD-linked genes function in synaptic processes such as cell adhesion, guidance/outgrowth, neurotransmission, signaling, degradation, transcription, and translation. A smaller fraction of ASD risk genes possessed unknown functions or “Other” non-synaptic functions. Examples of synaptically enriched ASD gene functions are modeled in Figure 5.

7. Conclusion
In conclusion, the broadened molecular landscape for ASD suggests that an integrated approach is required to understand functional pathways underlying ASD. An unbiased view of ASD risk gene datasets emphasizes the importance of overall synaptic networks for human cognition. Higher order functions require efficient information processing, and mutations in any synaptic component could lead to the range of impairments present in ASD. Future spatiotemporal mapping of ASD gene expression patterns may provide clues to how shared susceptibility genes give rise to different forms of ASD. Moreover, identification of new ASD-associated genes using advanced techniques like deep sequencing will increasingly sharpen our functional understanding of ASD synapse biology.
Fig. 5. Synaptic enrichment of ASD candidate genes. ASD risk genes are concentrated in numerous synaptic functions, including cell adhesion (examples: NLGN1, NLGN3, NLGN4X, NRXN1), synaptic scaffolds (examples: SHANK2-3), degradation (example: UBE3A), translation (example: FMR1), and neurotransmission (examples: CACNA1C, CACNA1G, GABRB3).

8. Acknowledgements

The authors would like to thank the other members of Mindspec, Inc. (Ajay Kumar, Cynthia Soderblom, Nicole Johnson, Rachna Wadhawan, Rainier Rodriguez, and Sue Spence), as well as the Simons Foundation. AutDB is licensed to the Simons Foundation as SFARI Gene.

9. References


Estimated prevalence rates of autism spectrum disorders (ASDs) have increased at an alarming rate over the past decade; current estimates stand as high as 1 in 110 persons in the population with a higher ratio of affected males to females. In addition to their emotional impact on the affected persons and their family members (in fact, the latter are often unrecognized unaffected â€œpatientsâ€ themselves), the economic and social impacts of ASDs on society are staggering. Persons with ASDs will need interdisciplinary approaches to complex treatment and life planning, including, but not limited to, special education, speech and language therapy, vocational skills training and rehabilitation, social skills training and cognitive remediation, in addition to pharmacotherapy. The current book highlights some of the recent research on nosology, etiology, and pathophysiology. Additionally, the book touches on the implications of new research for treatment and genetic counseling. Importantly, because the field is advancing rapidly, no book can be considered the final word or finished product; thus, the availability of open access rapid publication is a mechanism that will help to assure that readers remain current and up-to-date.

**How to reference**
In order to correctly reference this scholarly work, feel free to copy and paste the following: