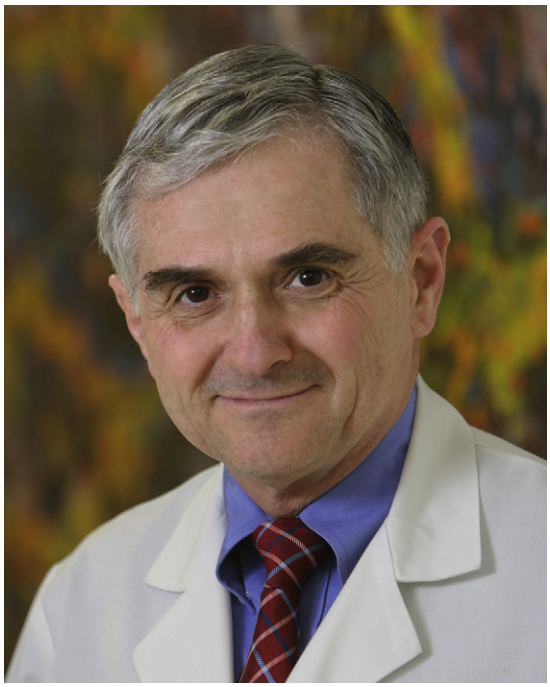
Allan Award Lecture: Rare Patients Leading to Epigenetics and Back to Genetics

Arthur L. Beaudet[1,*]



Arthur L. Beaudet

It is a great honor to receive the Allan Award today, and I want to thank Jim Lupski for his extremely kind remarks. I am sure that receiving this award will always be one of the highlights of my academic career.

I started doing human and mammalian genetic research at the NIH as a summer student in 1966. I returned to the NIH full time as a research associate in 1969, and both times, I worked in the laboratory of Tom Caskey in the larger overall unit headed by Marshall Nirenberg. Marshall was described in a recent edition of *Scentific American* as "The Forgotten Code Cracker." This was an extraordinary era for the NIH with the military draft for the Vietnam War funneling highly competitive graduates from U.S. medical schools into the NIH as a way of fulfilling military obligations, while at the same time receiving research training in an outstanding environment. Formalized NIH supported M.D./Ph.D. programs did not exist yet, or I am sure that I would have enrolled in one of them. In one small laboratory in the NIH in my first full year, Greg Milman, Joe Goldstein, and Ed Scolnick were present in addition to Tom Caskey. This was an extraordinarily stimulating environment for someone who had minimal previous exposure to research. All of these individuals had enormous impact on me and served as absolutely outstanding role models.

In 1971, I moved to Baylor College of Medicine in Houston with Tom Caskey and a year later initiated my own research program, and I have never been on the faculty anywhere else. As you can probably tell from Jim's introduction, the first 16 years of my research career were relatively lacking in focus. While I was jumping from lysosomal storage disease to urea cycle disorders and numerous other genetic disorders, Mike Brown and Joe Goldstein, my former benchmate, had won a Nobel Prize seemingly before my research career had gained any traction. Hopefully, I would be a late bloomer.

It was not until 1988, when we came to understand a patient with cystic fibrosis and uniparental disomy, that some semblance of focus on genomic imprinting and epigenetics took root. Soon thereafter, Allan Bradley arrived at Baylor College of Medicine, bringing with him the embryonic stem (ES) cell technology and homologous recombination that have just this month led to a Nobel Prize for Oliver Smithies, Martin Evans, and Mario Capecchi. Allan Bradley and the ES cell technology had an enormous transforming effect on genetics at Baylor College of Medicine. This further consolidated my focus on the use of mouse models to study the role of genomic imprinting and epigenetics in human disease.

I am sure that all of you know the general story of uniparental disomy (UPD) now, but we first described the phenomenon in a teenage girl with cystic fibrosis and short stature at the plenary session of this meeting in 1987. Ed Spence, who is here at the meeting, was a clinical fellow and did most of the lab work. This was the first description of a documented case of uniparental disomy in a human, but Eric Engel had suggested the potential for this phenomenon seven years earlier in 1980 in an article entitled "A new genetic concept: Uniparental disomy and its potential effect, isodisomy."[1] Obviously, we were very excited about this story, and we submitted it to *Science*, which promptly declined to send it out for review. Happily, the story ultimately found a home in the journal of this

[1]Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, BCM225, Houston, TX 77030, USA
*Correspondence: abeaudet@bcm.edu
DOI 10.1016/j.ajhg.2008.04.009.

society. We concluded with a comment that "uniparental disomy in an individual with a normal chromosome analysis is a novel mechanism for the occurrence of human genetic disease." I don't think that we fully appreciated the fact that this patient had both a genetic disorder, cystic fibrosis, and an epigenetic disorder, short stature analogous to the Russell-Silver syndrome. In an invited commentary, Dorothy Warburton, last year's Allan Award winner, stated the following: "It seems unlikely, then, that uniparental isodisomy will turn out to be anything but an interesting rarity. However, the demonstration of its existence is an extraordinary piece of human genetic sleuthing, which provokes the same admiration for the detectives and satisfaction in a carefully reasoned conclusion as does a good mystery novel." She correctly calculated that it would be very rare for uniparental disomy to lead to homozygosity for a recessive disorder, but I believe that neither we nor she appreciated at the time the potential for uniparental disomy to produce epigenetic phenotypes such as Prader-Willi and Angelman syndromes at moderately common frequencies. In fact, in the discussion of our paper, we said that "the short stature might be due to either embryonic chromosomal mosaicism or a second recessive genetic disorder on chromosome 7," failing to focus on the potential role of genomic imprinting, although we did mention the findings of Cattanach that both maternal and paternal chromosomes were required in mice for normal development. The word "imprinting" did not appear in our publication or in Dr. Warburton's commentary.

David Ledbetter was at Baylor College of Medicine during this interval, and he had described the interstitial deletions of chromosome 15q11-q13 giving rise to Prader-Willi syndrome some 7 years earlier.[2] David and I would frequently discuss the possible mechanisms to explain Prader-Willi syndrome in those patients who did not have deletions. This led to one of two missed opportunities that I look back on with a good-natured embarrassment. Given the cystic fibrosis UPD experience and our interest in Prader-Willi syndrome, David and I should have anticipated that non-deletion cases of PWS could be explained by maternal UPD for chromosome 15q11-q13. Instead, we left this discovery to Rob Nicholls and colleagues to be described two years later in 1989.[3]

We now know that most cases of Prader-Willi syndrome are caused by paternal deletions of 15q11-q13 and that most cases of Angelman syndrome are caused by maternal deletions of the same region. Now, some years later, it is perfectly obvious that deletion is a genetic form of PWS or AS while UPD is an epigenetic form of PWS or AS. If you could sequence the genome of a UPD patient, it would be normal and give no clue to the source of the phenotypic defect except the possibility of complete lack of heterozygosity at the nucleotide level for part or all of the chromosome. In terms of potential lessons for other disorders, it should be noted that almost all of these genetic and epigenetic cases of PWS and AS are de novo as contrasted to being inherited events. Over the years I have grown to emphasize this distinction with a genetic disease being an aberration of nucleotide sequence causing a disease phenotype in contrast to an epigenetic disease, which can be defined as an aberration in epigenotype (stable/heritable change in gene expression) causing a disease phenotype in the absence of nucleotide sequence aberration. Both can lead to the same biochemical and phenotypic result through altered expression of the same genes.

The Prader-Willi/Angelman region of chromosome 15 includes a bipartite imprinting center with a portion near the promoter for the *SNRPN* gene, which causes PWS when deleted. Another region slightly upstream from *SNRPN* causes AS when deleted. The *necdin* gene is paternally expressed. The *SNRPN* gene is a prominent player in this region, with paternal expression of a bicistronic protein coding transcript. In addition, the *SNRPN* transcript extends downstream to encompass a series of snoRNA genes and includes an antisense transcript for *UBE3A*. We now know that Angelman syndrome is caused by maternal deficiency for *UBE3A*, which functions both as an ubiquitin ligase and as a transcriptional coactivator. At the end of my talk, I will comment on the growing evidence that paternal deficiency for the HB-II85 snoRNA cluster causes Prader-Willi syndrome.

Another humbling experience in biology occurred in 1994 when we discovered that the E6-associated protein (E6-AP) was encoded by a gene in the PWS/AS region. Because the E6-AP was expressed biallelically in cultured cells, we did not consider it a strong candidate to be involved in the etiology of Angelman syndrome. One of the reviewers pointed out that this was not necessarily a safe conclusion. It was not until three years later that we and the group of Joe Wagstaff identified mutations in what was by then called the *UBE3A* gene as the cause of Angelman syndrome.[4,5] Had we simply performed mutation studies for this gene, the identity of the Angelman gene could have been uncovered three years earlier.

The main message that I would like to convey in this part of the talk is that the molecular bases for Angelman syndrome and PWS are extremely complex. I talked earlier about the deletion cases which are genetic and de novo in origin. In contrast, the UPD cases are epigenetic but are also de novo in origin. There are patients with imprinting defects caused by a small deletion in the imprinting center. These patients have a small genetic defect which causes a larger epigenetic defect. These imprinting-center deletions can be inherited or de novo in origin. Then there are patients with epigenetic defects where no nucleotide sequence change can be identified. These appear to be entirely epigenetic in nature, and they are virtually always of de novo origin. There is evidence that some forms of assisted reproductive technology, such as in vitro fertilization and intracytoplasmic sperm injection (ICSI), rarely can cause this form of Angelman syndrome. Finally, there are the patients with point mutations in *UBE3A* usually resulting in loss of function. These families identify *UBE3A* as the primary gene in AS. These mutations are entirely

genetic, but they can be inherited or de novo. The Angelman experience has taught us that epigenetic defects can give the same phenotype as a genetic defect and that there can be extensive heterogeneity involving the molecular basis of a single phenotype. This story of mixed genetics and epigenetics is similar for Prader-Willi syndrome, Beckwith-Wiedemann syndrome, and various forms of pseudohypoparathyroidism as caused by mutations in the *GNAS* complex. This experience with PWS and AS has led us to propose a mixed epigenetic and genetic and mixed de novo and inherited (MEGDI) model for these disorders.[6] In the case of Angelman syndrome, a single gene is primarily involved, with all possible mechanisms contributing to particular cases. One can easily imagine a MEGDI model for an oligogenic phenotype.

At about the same time that the *UBE3A* gene was implicated in Angelman syndrome, Ed Cook and colleagues published a report entitled "Autism or atypical autism in maternally but not paternally derived proximal 15q duplication."[7] This report shouted at us that we should investigate the role of the PWS/AS domain in autism. We identified a family from the Autism Genetic Resource Exchange (AGRE) collection with the type of duplication described by Cook, et al. and the data for a short tandem repeat in the PWS/AS region showed that two autistic sons inherited three alleles for the marker. Their father was normal and transmitted opposite alleles to his sons. The mother also had the duplication with two allele sizes on one chromosome and one allele size on the other. She transmitted the duplication chromosome to each of her sons. The inheritance of the duplication was confirmed by FISH studies. Pulsed field gel analysis using a methylation sensitive restriction enzyme and a probe from the PWS/AS domain for Southern blotting confirmed that the duplication was on the paternal chromosome in the mother and on the maternal chromosome in the autistic sons.

I would like to turn to autism in some greater detail. I don't need to define autism for this audience. It is the social and behavioral disabilities that distinguish autism. I believe it is useful to divide the autism patient population into two groups: Those with dysmorphic features and mental retardation and those who are nondysmorphic and higher functioning. This can be very important because some experts in the field will exclude the dysmorphic group from a study population while others will study a group comprised largely of dysmorphic cases. The concordance rate for monozygous twins is very high and the concordance rate for dizygous twins substantially lower. The sibling recurrence risk has been reported to be relatively low in the past, but there is some reexamination of this question with the possibility of a somewhat higher sibling recurrence risk with 5% perhaps being the lower end of estimates and 10% being the upper end. The overall sex ratio for autism is 4:1 male to female, but the sex ratio is substantially higher and perhaps in the range of 8:1 in the nondysmorphic high-functioning group. As I have just mentioned, there is evidence that genomic imprinting is important in cases where duplications of the Prader-Willi/Angelman domain are involved. The Rett gene, *MECP2,* with its effects on chromatin modification is implicated in some cases of autism. There is the intriguing question of whether autism is truly increasing in frequency.

Turning for a moment to the MZ/DZ twin story, the concordance in MZ twins is reported to be 60%–90% depending upon whether one uses a broader or narrower definition of the phenotype. The concordance in DZ twins is remarkably low. Perhaps it should be emphasized that these have been very small studies and that a much larger twin study is currently in progress. However, I want to specifically address the issue of a very high concordance in MZ twins but a low concordance in DZ twins, as this has been argued to perhaps be best explained by a large number of loci contributing to the causation of autism. I want to emphasize that if one has a phenotype caused by new mutation, and if the reproductive fitness of the affected individuals is low, one will observe a very high concordance of 100% in MZ twins and a low or very low concordance in DZ twins. Down syndrome caused by trisomy 21 is an excellent example of this principle. If one considers new mutation cases of achondroplasia or Rett syndrome in females, again there is a concordance of presumably 100% in MZ twins and a very low concordance in DZ twins. For virtually all of the forms of autism for which we identify genetic abnormalities, the expected concordance in MZ wins would be 100%. The concordance in DZ twins might be extremely low as in the case of de novo genomic deletions, or be higher as in the case of heritable single gene disorders such as fragile X syndrome. We could speculate about the possibility of concordance in MZ and DZ twins for de novo epigenetic defects such as UPD that I discussed as causing Angelman syndrome. If these arose prior to the timing of MZ twinning, and if the change were relatively stable, one could envision very high concordance in MZ twins and low concordance in DZ twins once again.

Heritability is often estimated based on concordance in MZ twins or on sibling recurrence risk. In the case of de novo events causing a phenotype, the MZ concordance will be very high, but the sibling recurrence risk may be very low. I think that it is clear that phenotypes such as trisomy 21 and deletion cases of PWS and AS can be highly heritable but not inherited if reproductive fitness is very low and new mutations or new epimutations are common. So how does this thinking fit with usual definitions of heritability? Heritability is traditionally defined as that proportion of the observed variation in a particular phenotype and in a particular study that can be attributed to the contribution of genotype (inheritance). By this definition, trisomy 21 and females with Rett syndrome would have a heritability of 1. Traditional definitions of heritability give little or no attention to epigenetics. Epimutations might be erased and reset in the next generation or they may be inherited in a semistable fashion, as has been shown in mice by the group of Emma Whitelaw.[8] Perhaps in the future we will need to distinguish genetic heritability from epigenetic heritability.

I would like to return now to the distinction between the two autism groups of dysmorphic with mental retardation and nondysmorphic with higher functioning. These are not absolute distinctions, but I believe that the majority of patients can be assigned to one or the other group. For the dysmorphic group, there is a high frequency of genetic abnormalities, and genomic deletions and duplications are now being recognized with much greater frequency.[9,10] Most of the genetic abnormalities causing dysmorphic autism involve autosomal abnormalities, and the male:female sex ratio is only slightly higher than 1:1,[11] although there is some male bias for genes on the X chromosome. For the nondysmorphic and high-functioning group, the frequency with which genetic abnormalities can be identified is substantially lower, at least at the present time. The sex ratio for this nondysmorphic and high-functioning group is more extreme, with perhaps approximately an 8:1 male-to-female ratio. Thus, we are left with a relatively large group of autism patients of unknown etiology, the majority of whom are male, nondysmorphic, and higher functioning.[12] It seems to me at this time that as much as 30%–50% of autism might be caused by point mutations and single-gene disorders (e.g., tuberous sclerosis and fragile X syndrome) and chromosomal abnormalities such as duplications of 15q11-q13, sex aneuploidy, and many de novo deletions and duplications. In the single-gene category, there is now an increasing number of genes encoding proteins that function at the synapse such as neuroligin 4 (*NLGN4X*) and *SHANK3* joining the likes of fragile X syndrome, Rett syndrome, and tuberous sclerosis. For the unknown group that might be 50%–60% of the patients, I believe that the etiology is almost completely unknown. One can imagine that the great majority of these individuals will turn out to have genetic disorders with many genes contributing to the phenotype in a single patient in what has been described as a multilocus epistatic model.[13] I would like to think that epigenetic abnormalities are a consideration. Environmental factors could play an important role.

The recognition that many patients with autism have de novo genomic deletions or duplications is not a new finding, although the magnitude of this etiology has greatly increased in awareness during the past year. As nicely reviewed by Vorstman, et al. in 2006,[14] deletions and duplications have been observed for virtually every human chromosome. Duplications of chromosome 15q11-q13 are the most frequent findings, with the telomeres of chromosomes 2q and 22q also being prominently involved. During the past year, two publications have greatly increased the awareness of the potential for a much higher frequency of these abnormalities contributing to autism. A paper by Jacquemont et al.[9] found deletions or duplications in 27% of 29 patients with syndromic autism. Sebat, et al.[10] found de novo deletions or duplications in 10% of simplex cases and 1% of controls. These papers have led to a substantial upward revision of the percentage of cases of autism with de novo genomic abnormalities. In addition, we know that the screening for copy-number abnormalities to date is relatively crude and may only have detected a small fraction of the total of such mutations. The photographs in the supplemental materials of the publication by Jacquemont et al.[9] typify the dysmorphic phenotypes seen with autism and mental retardation. These investigators found 28% of patients having de novo mutations, as contrasted to 10% for Sebat, et al.[10] The patients with genomic abnormalities have mostly new mutations, and the male:-female ratio is closer to 1:1 than for the rest of the autism population. Because heterozygous deletions of genes can cause autism (e.g., *SHANK3*), it seems ensured that there will be point mutations causing autism for many of these same genes. The probability that de novo point mutations might also be very important in the etiology of autism is suggested by the evidence of an advanced-paternal-age effect in autism. Earlier reports appear to be substantiated by a more recent study indicating an approximately 6-fold increase in the risk of autism for men over 40 years of age, and one could even imagine a unique mutation contributing to a significant fraction of autism, as occurs for achondroplasia and progeria.

Attempting to specify what proportion of autism patients have a particular etiology is heavily dependent on whether patients with dysmorphic features and mental retardation are included or excluded from consideration. In my view, the various collections of autism patients differ dramatically in their selection bias. In a genetic clinic series and in the type of patients reported by Jacquemont et al.,[9] there is an under-ascertainment of patients who are nondysmorphic and higher functioning. On the other hand, in the child psychiatry clinic or in a series such as the AGRE collection, there is an under-ascertainment and even intentional exclusion of patients with dysmorphic features and mental retardation. This leads to a very high frequency of genetic abnormalities in the dysmorphic series and a relatively lower detection of abnormalities in the nondysmorphic series.

For some time now, my laboratory has been in engaged in a search for epigenetic causes of autism. We have focused on analysis of autopsy brain because of the possibility that epigenetic abnormalities might be brain specific. Although we have found one autism brain sample with a DNA methylation abnormality at the 5′ end of the *UBE3A* gene,[6] we have not found epigenetic abnormalities at a substantial frequency up to the present. However, it is only recently that we have put the methods in place to systematically search across the genome for epigenetic abnormalities in autism brain as compared to control brain. Chromatin immunoprecipitation analyzed by arrays (ChIP-chip) is proving to be a powerful tool with which we can detect known abnormalities of chromatin structure in PWS or AS brain compared to control brain. Similarly, the differences of DNA methylation between PWS or AS and control brain can be demonstrated using genome-wide methods. We consider this to be a proof of feasibility indicating that the methodology is in hand to discover striking epigenetic defects in brain from autism, schizophrenia, or bipolar

disorder patients if such abnormalities occur and represent a significant fraction of the pathogenesis.

In my title for this talk, I referred to coming back to genetics. Routine clinical laboratory studies using array CGH were performed on a patient who was thought to have Prader-Willi syndrome, but previous laboratory studies were normal, including deletion analysis and methylation analysis. The patient was described in much greater detail by Dr. Trilochan Sahoo this morning. Array CGH using a BAC array showed a relatively subtle loss of copy number for one clone in the PWS/AS region. FISH studies with this BAC showed the possibility of a weaker signal for one chromosome, but there was not a complete deletion of the BAC. Methylation studies were confirmed to be normal. When DNA from this patient was analyzed with a chromosome 15 custom 44K Agilent array, there was precise definition of a deletion that removes the entire HBII-85 snoRNA cluster and about half of the HBII-52 snoRNA cluster (our unpublished data). From previous translocation and deletion cases, there had been moderately compelling evidence that the PWS phenotype might be caused by paternal deficiency for the HBII-85 cluster of snoRNA genes. We believe that this single patient with a unique genetic lesion greatly adds to the evidence that PWS is caused by paternal deficiency for the HBII-85 cluster of snoRNAs. Thus, as previously suggested by Uta Francke and her colleagues, the evidence is now stronger than ever that the PWS phenotype is caused by paternal deficiency for the HBII-85 snoRNA cluster. As of the moment, there is no solid understanding of the possible function of the HBII-85 snoRNAs, and this remains an important subject for future investigation.

I would like to conclude by thanking the numerous people who have worked in my laboratory or collaborated with us over the last three and a half decades. I might especially mention William O'Brien, who collaborated with me in many of the earlier years, and Brendan Lee and Huda Zoghbi, both of whom spent time in my laboratory as K awardees. With junior colleagues such as these, one is bound to enjoy some successes. Again, I would like to thank Jim Lupski for his most kind introduction and ongoing interactions.

References

1. Engel, E. (1980). A new genetic concept: Uniparental disomy and its potential effect, isodisomy. Am. J. Med. Genet. *6*, 137–143.
2. Ledbetter, D.H., Riccardi, V.M., Airhart, S.D., Strobel, R.J., Keenan, B.S., and Crawford, J.D. (1981). Deletions of chromosome 15 as a cause of the Prader-Willi syndrome. N. Engl. J. Med. *304*, 325–328.
3. Nicholls, R.D., Knoll, J.H.M., Bulter, M.G., Karam, S., and Lalande, M. (1989). Genetic imprinting suggested by maternal heterodisomy in non-deletion Prader-Willi syndrome. Nature *342*, 281–285.
4. Kishino, T., Lalande, M., and Wagstaff, J. (1997). *UBE3*A/E6-AP mutations cause Angelman syndrome. Nat. Genet. *15*, 70–73.
5. Matsuura, T., Sutcliffe, J.S., Fang, P., Galjaard, R.-J., Jiang, Y.-H., Benton, C.S., Rommens, J.M., and Beaudet, A.L. (1997). De novo truncating mutations in E6-AP ubiquitin-protein ligase gene (*UBE3A*) in Angelman syndrome. Nat. Genet. *15*, 74–77.
6. Jiang, Y.H., Sahoo, T., Michaelis, R.C., Bercovich, D., Bressler, J., Kashork, C.D., Liu, Q., Shaffer, L.G., Schroer, R.J., Stockton, D.W., et al. (2004). A mixed epigenetic/genetic model for oligogenic inheritance of autism with a limited role for UBE3A. Am. J. Med. Genet. *131A*, 1–10.
7. Cook, E.H.J., Lindgren, V., Leventhal, B.L., Courchesne, R., Lincoln, A., Shulman, C., Lord, C., and Courchesne, E. (1997). Autism or atypical autism in maternally but not paternally derived proximal 15q duplication. Am. J. Hum. Genet. *60*, 928–934.
8. Chong, S., Youngson, N.A., and Whitelaw, E. (2007). Heritable germline epimutation is not the same as transgenerational epigenetic inheritance. Nat. Genet. *39*, 574–575.
9. Jacquemont, M.-L., Sanlaville, D., Redon, R., Raoul, O., Cormier-Daire, V., Lyonnet, S., Amiel, J., Le, M.M., Heron, D., De Blois, M.-C., et al. (2006). Array-based comparative genomic hybridization identifies high frequency of cryptic chromosomal rearrangements in patients with syndromic autism spectrum disorders. J. Med. Genet. *43*, 843–849.
10. Sebat, J., Lakshmi, B., Malhotra, D., Troge, J., Lese-Martin, C., Walsh, T., Yamrom, B., Yamrom, B., Yoon, S., Krasnitz, A., et al. (2007). Strong association of de novo copy number mutations with autism. Science *316*, 445–449.
11. Miles, J.H., and Hillman, R.E. (2000). Value of a clinical morphology examination in autism. Am. J. Med. Genet. *91*, 245–253.
12. Beaudet, A.L. (2007). Autism: highly heritable but not inherited. Nat. Med. *13*, 534–536.
13. Bailes, J.A., and Monaco, A.P. (2007). Genetic Basis of Autism. In Understanding Autism: From Basic Neuroscience to Treatment, S.O. Moldin and J.L.R. Rubenstein, eds. (Boca Raton, FL: CRC Press), pp. 49–74.
14. Vorstman, J.A.S., Staal, W.G., van Daalen, E., van Engeland, H., Hochstenbach, P.F., and Franke, L. (2006). Identification of novel autism candidate regions through analysis of reported cytogenetic abnormalities associated with autism. Mol. Psych. *11*, 18–28.