Project no.512158

AUTISM MOLGEN

Using European and International Populations to identify Autism Susceptibility loci

Specific Targeted Research Project

Priority 1 Life Sciences, Genomics and Biotechnology for Health

Final Activity report

Period covered: from 1/10/05 to 31/3/09 Date of preparation: 20/7/09

Start date of project: 1/10/05 Duration: 42 months

Project coordinator name Anthony Bailey
Project coordinator organisation name University of Oxford

Section 1. Project summary and objectives

The primary objective of the EU Autism MOLGEN project was to identify susceptibility alleles for autism spectrum disorders (ASDs). Therefore, the consortium aimed to pool data and laboratory resources, as well as exploit genetically isolated European populations, such as those from Finland and the north of Holland. These resources would then be used to examine ASD candidate loci using methods including linkage, parent-of-origin analysis, association and mutation screening. To achieve these aims, the project was divided into six work-packages covering the following key themes; family collection, linkage studies, candidate gene analyses, innovation activities and project management. These where initially envisaged as taking place over a period of 36 months. However, with the arrival of new genotyping technologies enabling a more detailed search for susceptibility loci, an extension to 42 months was successfully applied for during the initial reporting period.

Work-package 1 covered the collection of new autism families, both singleton and multiplex. In recent years, the increasing awareness of the role of rare variants, including both single nucleotide polymorphisms (SNPs) and copy number variants (CNVs), has made the acquirement of new families of vital importance. In addition, the complex nature of the genetic susceptibility to ASD, makes increasing cohort sizes of great importance in order to identify common variants of weak to moderate effect, or rare variants of large effect. Therefore, the project aimed to collect an additional 100 multiplex and 250 singleton families.

Work-package 2 focused primarily on narrowing the peaks of linkage in those regions previously identified by members of the consortium. Despite technological progression in genotyping methodologies now allowing large scale association studies, including at the genome-wide level, linkage studies continue to remain of value to identify susceptibility loci. By trying to refine these regions, we aimed to help in the identification of the specific causal loci underlying them and reveal new candidate genes for more focused association studies. Therefore, it was decided to undertake a meta-analysis of linkage

using three available cohorts, IMGSAC, PARIS and Finnish combined with additional data generated by the AGP. This work would be supplemented by parent-of-origin analyses and the use of two relatively isolated populations, Finnish and Northern Dutch, to examine the regions of linkage for extended risk haplotypes. In order to take advantage of progress in genotyping methodologies, we revised our original work plan to use higher throughput technology (Illumina).

Work-packages 3 and 4 together comprised the candidate gene analysis portion of the project. Both work-packages originally contained separate genotyping elements. Specifically, work-package 3 examined candidate genes in regions of linkage, while work-package 4 investigated candidate genes not necessarily within these loci. However, as with work-package 2, these portions of the project were adapted in light of new genotyping technologies. To make use of Illumina GoldenGate® genotyping facilities in Oxford, the project was revised to centralise the genotyping to this site. In addition, the experiment was designed to include seven genes covering elements of both workpackages, *NOSTRIN*, *RELN*, *GRIK2*, *PRKCB1*, *SLC6A4*, *SHANK3* and *ASMT*. This more focussed candidate gene based approach has allowed us to obtain greater coverage of variation in genes specifically implicated in ASD by previous studies from our consortium members. In addition to the centralised genotyping of the seven genes listed, additional mutation screening and association studies were undertaken by the consortium at various sites.

Work-packages 5 and 6 involved innovation and project management activities, respectively, to allow direction of the consortium, dissemination of results and development of any intellectual property that it should produce.

A detailed break down of the objectives of the project is given in Table 1.1.

Table 1.1 Summary of the Autism MOLGEN Project objectives

Work-package	Task	Description
1	1.1	Ascertainment and assessment of at least 100 affected relative pairs.
	1.2	Ascertainment and assessment of at least 250 singleton families.
2	2.1	Perform a meta-analysis of linkage in the top six susceptibility regions (2q, 3q, 6q, 7q, 16p and 17q) using affected relative pairs (ARP) from the IMGSAC, PARIS and Finnish multiplex samples.
	2.2	Analyse fine mapping data on chromosomes 6q and 7q for parent-of-origin linkage effects in IMGSAC, PARIS, and Finnish multiplex samples.
	2.3	Assessment of methylation/imprinting status of candidate genes.
	2.4	Test the top susceptibility regions in Finnish singleton trios and Friesland (Netherlands) singleton trios to search for extended haplotypes that may further refine the critical regions.
3	3.1	Screen each candidate genes in 30-50 probands from sharing multiplex families by DHPLC, and identify all variants by DNA sequencing.
	3.2	Identify single nucleotide polymorphisms (SNPs) from the public databases that tag the haplotypic diversity across these genes in the multiplex and singleton samples.
	3.3	Screen new variants and haplotype tag SNPs (htSNPs) for association in both multiplex and singleton families from IMGSAC, PARIS, Finland, Friesland, Germany and Italy.
4	4.1	Test PARIS and Finnish multiplex samples for identified mutations in *RELN*, *LAMB1* and *CGEF2*.
	4.2	Test high density SNPs across *RELN*, *PRKCB1* and *NOSTRIN*.
	4.3	Test IMGSAC and Finnish multiplex samples for mutations and association in *GluR6* (*GRIK2*), *ASMT*, *NLGN3* and *NLGN4*.
	4.4	LD analysis of *SLC6A4* in IMGSAC, PARIS and Finnish families.
	4.5	Test mutations/SNP associations confirmed in multiplex families.
5	5.1	Intellectual property protection.
	5.2	Dissemination.
6	6.1	Coordination activities.
	6.2	Reports and reviews.

Section 2. Results

Over the lifespan of the project, various aspects of the work-packages have been modified in light of both advances in technology and the autism genetics field in general. Therefore, we have met the majority of our original aims, but in some cases have modified our work to take advantage of the latest developments. A summary of the achievements in each work-package over the 42 months of the project are given below.

2.1 Work-package 1 - Phenotypic assessment of multiplex and singleton families

The aim of work-package 1 was the collection of 100 multiplex and 250 singleton families and involved the 13 clinical sites of the consortium, in addition to a collaborating group based in Chicago. At the end of the project the final family numbers collected were 102 multiplex and 372 singleton. We have therefore achieved our target in terms of multiplex family collection and have considerable exceeded the number of singleton families we aimed to collect.

2.2 Work-package 2 - Meta-analysis of linkage and search for extended haplotypes

The primary focus of work-package 2, Task 2.1, was the refinement of ASD linkage regions. The majority of this work was performed using an Illumina GoldenGate® 384 SNP genotyping array. Six regions previously reported to show linkage to ASD, 2q, 3q, 6q, 7q, 16p, and 17q, were investigated. SNPs were chosen by a haplotype tagging approach and genotyped in our IMGSAC, PARIS and Finnish cohorts. This data was then combined with genotypes for the same SNPs for additional IMGSAC and PARIS families that had been generated by the Autism Genome Project (AGP) consortium as part of a genome wide linkage screen. A meta-analysis of linkage was performed, with the strongest signal observed on chromosome 2q (MLOD = 1.89, rs2885116) which approached suggestive levels of linkage (LOD > 1.9). No other significant results were

observed at the other loci examined. In addition to this main linkage analysis, X chromosome inactivation (XCI) was examined in a set of IMGSAC, PARIS and Finnish individuals. No significant evidence of skewed X chromosome inactivation was found. However it was noted that in some families with mutations in X-linked genes there was significant skewing. This indicated that skewed genes could indicate underlying mutations and so a subset of families showing skewed XCI was used for linkage analysis. This resulted in evidence of a modest increase in allele sharing on Xq27-Xq28.

In Task 2.2 it had originally been proposed to examine parent-of-origin effects from chromosomes 6q and 7q using the linkage genotyping data. This analysis was extended to include all six loci from the linkage meta-analysis in order to make maximum use of the data available. While some parental effects were observed, none approached suggestive levels of linkage and so did not warrant further investigation.

Due to the lack of strong positive results in Tasks 2.1 and 2.2, the decision was made to not proceed with Task 2.3. Task 2.3 had been envisaged as building upon any positive linkage results from the previous two tasks by examining methylation patterns for candidate genes in the linked regions. However, due to the lack of such positive signals, it was felt that resources would be better directed to other tasks within the project.

Finally, Task 2.4 was to examine our more genetically isolated cohorts of Finnish and Northern Dutch samples for extended haplotypes in the regions genotyped as part of Task 2.1. However, we found no difference in linkage disequilibrium between these cohorts and that of the HapMap project. Therefore, we decided to use the data to perform a test of association. This resulted in five significant associations surviving correction for multiple testing being identified. Three associations were in the Finnish cohort, one of which was in *MKL2* and the remaining two on the X chromosome. In the Northern Dutch cohort the associations were found in *SND1* and on the X chromosome. None of these associations replicated in the IMGSAC cohort and so could indicate population specific ASD risk factors.

2.3 Work-package 3 - Candidate gene analysis in regions of linkage

Work-package 3 was the first of two work-packages within the project concerning the investigation of candidate genes for ASD. Work-package 3 focussed specifically on those candidate genes located within previously reported regions of linkage.

For Task 3.1 30 genes across a variety of linkage loci have been screened for variants in ASD samples. Screening focussed on the exons, flanking intronic sequence and promoter regions. The 30 genes examined were ***AA-NAT***, *ACSL4*, ***ASMT***, *ATP6AP1*, *DLG3*, *GPR50*, ***GRIK2***, *HCFC1*, *IL1RAPL2*, *KCNJ3*, ***MTNR1A***, ***MTNR1B***, *NLGN1*, ***NLGN2***, ***NLGN3***, ***NLGN4***, *NOSTRIN*, *NR4A2*, ***NRXN1***, *RP56KA6*, ***RPL10***, *RPRM*, ***SEZ6L2***, ***SHANK1***, ***SHANK2***, ***SHANK3***, *STK23*, *UPP2*, *ZNF11* and *ZNF533*. Of these, variants of interest which may influence ASD susceptibility were identified in 14 genes, identified in bold in the list above.

In addition to SNP screening, some examination for larger copy number variants (CNVs) has also been performed, due to the increasing volume of published research indicating that this class of variant is of importance in ASD susceptibility. A CNV of putative importance identified as part of this study was found in the *IMMP2L-DOCK4* region of chromosome 7. Also, due to the variants, including CNVs, identified in our initial screen of *SHANK3* and the 22q13 region, this locus was specifically examined for CNVs in a subset of the IMGSAC cohort, however no additional CNVs were found.

Tasks 3.2 and 3.3 together formed a large portion of the genotyping of work-package 3. In Task 3.2 haplotype tagging SNPs were identified for *GRIK2*, *SHANK3* and *ASMT* using the publically available HapMap data. As part of the consortium's revised scientific plan, genotyping and analysis of these SNPs in the IMGSAC, PARIS and Finnish cohorts for Task 3.3 was centralised in Oxford and performed using an Illumina GoldenGate® 384 SNP array to make use of this new high-throughput genotyping technology. However, no associations surviving multiple testing were found to any of the SNPs in our meta-analysis. The genotyping array used for this experiment also contained SNPs as part

of work-package 4. Therefore, the five SNPs with the strongest associations, including two SNPs in *GRIK2* (rs2518256 and rs2518261) for work-package 3, were taken forward for replication in the Northern Dutch cohort and additional IMGSAC samples. Again, no association surviving correction for multiple testing was found to the two *GRIK2* SNPs in either the replication samples or the total combined set of samples.

In addition to the centralised genotyping performed in task 3.3, we have also investigated additional genes located in linkage regions for association with ASD. Fourteen genes in the 7q linkage region were investigated in German families (*SEMA3C*, *PCL0*, *FZD1*, *TAC1*, *NPTX2*, *ACHE*, *SERPINE1*, *AP1S1*, *CLDN15*, *SMO*, *SEC8L1*, *FLJ32786*, *YEA*, *AKRIB1*) with borderline significance obtained in six (*NPTX2*, *SERPINE1*, *SMO*, *SEC8L1*, *FLJ32786*, *AKRIB1*). In two separate studies, *KIF1A* and *RPL10* were also examined for association in the German cohort, with only weak evidence of association being found to one SNP in *KIF1A*. Finally, *EN2* was examined using samples from the IMGSAC cohort in a case-control study, but the results did not support association to ASD.

2.4 Work-package 4 - Testing of previously identified candidate genes

The final experimental work-package focussed on examining ASD candidate genes not specifically located within reported regions of linkage to ASD. Due to changes as part of the revised scientific plan for the project, particularly the use of an Illumina GoldenGate® 384 SNP genotyping array, elements of Tasks 4.1, 4.2 and 4.3 were combined. Haplotype tagging SNPs were identified for the genes *RELN*, *PRKCB1*, *NOSTRIN*, *SLC6A4*, *GRIK2* and *ASMT*, the latter two also analysed as part of work-package 3. Genotyping was then performed using an Illumina GoldenGate® 384 SNP array. No associations surviving correction for multiple testing were identified. However, it was decided to further investigate the five strongest associations (*GRIK2* – rs2518256, rs2518261; *PRKCB1* – rs9925126, rs11074601; *RELN* – rs362780) in additional IMGSAC and Northern Dutch samples as part of Task 4.5. No significant associations were found in the replication

samples, but when the data from all genotyped samples was pooled, the association to the *RELN* SNP rs362780 was found to increase to $P = 0.001$, although this would not survive correction for multiple testing.

Additional genotyping in five genes was performed in Task 4.2. Due to initial results implicating a role for Leucine Rich Repeat (LRR) genes in ASD susceptibility from collaborators in the AGP, haplotype tagging SNPs were used to investigate association in *LRRTM1*, *LRRTM3*, *LRRN1* and *LRRN3* in IMGSAC and Northern Dutch samples. Association surviving correction for multiple testing was identified in three of the genes. The strongest associations were for SNPs within *LRRTM3* and *LRRN3*, with the latter association being in the Northern Dutch cohort. Weaker evidence of association was also found for SNPs in *LRRN1*. In a separate experiment, two SNPs in *MET* previously reported as having shown association to ASD were also genotyped, however, no association could be found.

In Task 4.3, a further four genes, *SLC1A1*, *SLC1A2*, *JMJD2C* and *LDLRAD3*, were investigated for association to ASD in the Finnish cohort. The strongest signal identified was in *JMJD2C*, which while not surviving correction for multiple testing, was significant after permutation tests and therefore still of interest.

In addition to these association studies, several genes were also screened for sequence variants as part of Task 4.3. Three neuroligin genes, *NLGN3*, *NLGN4X* and *NLGN4Y*, were screened with only one synonymous mutation being identified. Association analysis of these genes also revealed only a very weak association in the region of *NLGN3* and *NLGN4*. A fourth gene, *ASMT*, was also screened, revealing several variants including a splice-site variant, non-synonymous variants and a stop mutation, the latter present in a single ASD family, but not in controls. Some evidence of association was found to two *ASMT* SNPs, rs4446909 and rs5989681, which were also associated with *ASMT* expression levels. However, the final conclusion was that there is no clear correlation between *ASMT* variants and ASD risk.

In Task 4.4, association was also examined for *SLC6A4*, but no significant results were identified. We also attempted to replicate previously reported associations of 3 SNPs in *SLC25A12* and *CMYA3* as part of Task 4.5, but again no evidence of association was found.

2.5 Work-package 5 - Innovation activities

During the course of the project, no discoveries requiring intellectual property protection, Task 5.1, were made. For Task 5.2, we have increasingly disseminated information related to the work of the project via a variety of avenues as the project has continued. These have included over 50 seminars, presentations, press releases, posters and workshops, in addition to over 40 scientific papers either published, in press or in preparation which include work performed with funding from this project.

2.6 Project management activities

During the project, ties have been maintained between all partners. In particular, this has been fostered by the annual meetings in Oxford in which results have been shared and general discussion about experiments held. Contact has also been maintained between individual groups, including visits between partners, leading to a number of experiments carried out at joint sites. Due to the nature of this strong collaboration, it is envisaged that the partnerships formed will continue on beyond the lifespan of the project itself.

Three periodic reports have been prepared for the project. Two, each covering one of the first two years of the project, have been accepted by the EU. The third, covering the final 18 months, has been submitted. A review of the second reporting period was performed by Professor Antonio Persico in April 2007, for which we received very favourable comments.

Section 3. Conclusions

During the course of this project we have met the key objectives in each work-package. Our collection of 102 multiplex and 372 singleton families will help achieve the additional sample numbers needed to dissect the genetic architecture of the ASDs and so will provide a valuable resource for future studies.

The molecular genetics portion of the project has also provided important results. This work was divided into linkage studies, association studies and mutation screens. Recent advances in genotyping technology and subsequent decreases in cost have allowed the advent of whole genome association studies. Despite this, linkage studies can still be of value in helping to identify and refine critical loci and thus identify putative candidate genes for further study. Towards this end, we attempted to refine several linkage peaks previously identified by our members. Disappointingly, LOD scores approaching suggestive evidence of linkage were only obtained for chromosome 2q. The lack of linkage signals in our analysis could be due to increased heterogeneity in our combined cohort. However, the positive result for chromosome 2q further confirms the importance of this region in ASD susceptibility. Also of interest is that the subsequent TDT analysis of the SNPs genotyped as part of the linkage study in the Finnish and Northern Dutch cohorts identified significant association in two genes, *SND1* and *MKL2*. These associations survived correction for multiple testing and so represent interesting potential new candidate genes for future study.

In addition to the association study performed using the genotype results from the linkage study, we also examined a further 37 genes for association to ASD. Of these, several nominal associations have been identified, but only a handful survived correction for multiple testing. This is result is not surprising, as for even the most well studied candidate genes in the scientific literature, such as *RELN*, multiple studies showing both association and lack of association have been published. Therefore, while for the majority of the genes we have investigated, we cannot conclude that our data supports their role in ASD susceptibility, neither would it be appropriate to categorically reject them as

unimportant in these disorders. However, we have amassed evidence for the role of several genes in ASD susceptibility. Key amongst these are *LRRTM3*, *LRRN1*, *LRRN3*, *SND1* and *MKL2*, each of which we have found association to, even at levels corrected for multiple testing. These genes present interesting avenues for future research. In addition, while our associations for genes such as *JMJD2C*, *RELN*, and *GRIK2* may not survive stringent correction for multiple testing, the associations observed are of interest and can be taken as part of a cumulative case for their involvement in ASDs. This is particularly the case for *RELN*, which has been implicated in multiple studies.

The final molecular genetics aspect of our work has been mutation screening in candidate genes. This type of investigation is becoming of particular relevance in ASD genetics due to the increasing numbers of reports implicating rare mutations of large effect influencing the development of these disorders, such as in the case of the neuroligins. We have identified rare variants of possible functional effect, including non-synonymous, splice-site and stop codon mutations, both in genes associated with ASD and in other plausible candidates. Such variants will serve as important starting points for future functional studies. In addition to such small variants, the importance of larger CNVs in ASD is also increasingly recognised. Although our genotyping data has not allowed investigation for such variants, as it has in more dense SNP genotyping studies, we have performed some focussed CNV identification studies. This has resulted in discoveries such as the CNV at the *IMMP2L-DOCK4* locus which have a putative functional effect.

Therefore, this project has met its stated objectives and helped identify ASD susceptibility alleles, contributed to the evidence for previously recognised candidate genes and loci, generated vital new sample collections and identified avenues for possible future research. The publication of our research results in high quality scientific journals also means that not only the members of this consortium, but also the wider research community, will benefit from the Autism MOLGEN Project.