A transcription map of the 6p22.3 reading disability locus identifying candidate genes

Eric R Londin1, Haiying Meng2 and Jeffrey R Gruen*2

Address: 1Graduate Program in Genetics, State University of New York at Stony Brook, NY, USA and 2Yale Child Health Research Center, Department of Pediatrics, Yale University School of Medicine, New Haven, CT, USA

Email: Eric R Londin - Eric.Londin@stonybrook.edu; Haiying Meng - Haiying.Meng@yale.edu; Jeffrey R Gruen* - Jeffrey.Gruen@yale.edu

* Corresponding author

Published: 30 June 2003
BMC Genomics 2003, 4:25
Received: 22 April 2003
Accepted: 30 June 2003
This article is available from: http://www.biomedcentral.com/1471-2164/4/25
© 2003 Londin et al; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article’s original URL.

Abstract

Background: Reading disability (RD) is a common syndrome with a large genetic component. Chromosome 6 has been identified in several linkage studies as playing a significant role. A more recent study identified a peak of transmission disequilibrium to marker JA04 (G72384) on chromosome 6p22.3, suggesting that a gene is located near this marker.

Results: In silico cloning was used to identify possible candidate genes located near the JA04 marker. The 2 million base pairs of sequence surrounding JA04 was downloaded and searched against the dbEST database to identify ESTs. In total, 623 ESTs from 80 different tissues were identified and assembled into 153 putative coding regions from 19 genes and 2 pseudogenes encoded near JA04. The identified genes were tested for their tissue specific expression by RT-PCR.

Conclusions: In total, five possible candidate genes for RD and other diseases mapping to this region were identified.

Background

Reading disability (RD), or dyslexia, is a common syndrome with a significant genetic component. Genetic linkage studies have identified five potential RD loci located on chromosomes 1 [1,2], 2 [3,4], 6 [5–9], 15 [7,10] and 18 [11]. The linkage to 6p has been the most often reproduced in independent samples with five studies showing peaks of linkage to different regional marker sets [5–9]. We recently constructed a BAC/PAC contig of the 6p RD locus [12] and identified the precise location and order of 29 short tandem repeat (STR) markers spanning this region [13]. A subsequent study with this new marker panel identified a peak of transmission disequilibrium at marker JA04 (G72384) [14].

We searched the expressed sequence tag (EST) database, dbEST http://www.ncbi.nlm.nih.gov/dbEST/index.html, with the genomic sequence corresponding to the peak of transmission disequilibrium at marker JA04. ESTs are partial and usually incomplete cDNA sequences prepared from various tissues. Presently, nearly four million ESTs (dbEST release 04/19/02) are catalogued in dbEST representing about 80% or more of all human genes with at least one representative entry [15]. While not every gene is accounted for in dbEST, computer database searching,
also known as in silico cloning, can identify new genes without actual physically manipulating DNA. These types of analyses can also characterize intron-exon boundaries, splice variants, tissue specific expression levels, and gene homologies [16]. Furthermore, clustering ESTs together to form a contiguous sequence can predict putative open reading frame (ORFs). The first map of the human genome, which contained over 30,000 genes, was generated by mapping EST clusters to human-hamster radiation hybrid cell lines [17]. Despite their wide-ranging utility, ESTs have two inherent drawbacks: (1) they are based on single sequence reads making them vulnerable to sequencing errors, and (2) they are generated from cDNA libraries that may contain unexpressed or incompletely spliced sequences derived from heteronuclear RNA or other artifacts. The clustering of ESTs to form ORFs could contain both expressed and unexpressed sequences and should be treated with caution. Fortunately, the high redundancy of entries in dbEST permits the alignment of multiple ESTs for most genes thus diminishing the effects of these drawbacks.

To identify candidate genes for RD and other disorders mapping to this region we downloaded and searched the two million base pairs of genomic sequence surrounding the peak of transmission disequilibrium. In addition to RD, risk loci for Behçet’s disease [18], inflammatory bowel disease (IBD3) [19,20], hypotrichosis simplex (HSS) [21], insulin dependent type 1 diabetes mellitus [22], attention deficit hyperactivity disorder (ADHD) [23] and schizophrenia [24] have all been assigned to this general chromosomal location by genetic linkage analysis. Using in silico cloning we identified a total of 19 genes and 2 pseudogenes and mapped their precise physical location and direction of transcription. The expression pattern of each gene was characterized by examining the number of ESTs identified from various tissues as well as by qualitative RT-PCR with RNA from 20 different human tissues. This study also allowed us to test the usefulness of in silico cloning to identify and map new genes in a focussed region of the genome.

**Results**

Using the blastc13 server at NCBI, we performed in silico cloning studies of the 6p RD locus to identify coding regions. In total, 623 ESTs from 80 different tissues were identified and aligned to 2 Mb of genomic sequence. These searches captured 157 putative coding regions from 19 genes and 2 pseudogenes concentrated in the central 1200 Kb shown in detail in Figure 1 with base pair 1 starting at the 5-prime end of FLJ12671 and ending 1 Kb centromeric to the 3-prime end of RPS10. Short tandem repeat marker JA04, which identified the peak of transmission disequilibrium for RD phenotypes in previous studies [14], is at 540 Kb. The most telomeric 200 Kb and the most centromeric 600 Kb of genomic sequence are void of coding regions. Intergenic distances range from less than 1 Kb (KIAA0319 and TRAF) to 110 Kb (HNRPAP1 and P24). Cytokeratin 8, transcribed telomere to centromere, is located in the intron between exons 1 and 2 of KIAA0319 (transcribed centromere to telomere). Table 1 lists the genes identified in Figure 1, their size in Kb, the NCBI accession number for the corresponding mRNA or cDNA, number of exons, genomic mapping source, and putative function.

Thirteen genes were previously mapped to this region (RU2AS, MR9S2L, GPLD1, SSADH, KIAA0319, TRAF, HT012, FLJ12619, Geminin, KIAA0386, Cyclophilin A, CMAH, and NUP50) and are present on the NCBI accession view (NT_017021) of 6p21.3-22. Our in silico studies identified an additional six genes (RPS10, FLJ12671, UBE2D2, AP3, Cytokeratin 8, and P24) and two pseudogenes (HNRPAP1 and ASSP2) not represented on NT_017021. In addition we identified a putative fifth exon for P24 from hypothalamus (BG715502) and pineal gland (AA363698) cDNA libraries, located between exons 2 and 3 in the 5-prime untranslated region and flanked by non-traditional splice donor and acceptor bases [25]. The novel exon was not present in the full-length mRNA sequence (AF418980).

The functions of the newly identified genes were inferred by their similarity to other known genes identified in the BLAST searches. Ribosomal protein L21 (RPS10) is 95% identical to the genomic and mRNA sequences of a known ribosomal gene. RPS10 is one of the many proteins that make up the ribosome macromolecule. Other highly homologous genes of RPS10 are RPS5, RPS9, RPS29, RPL5, RPL27a, and RPL28 [26]. FLJ12671 is a hypothetical gene with an unknown function identified by the NEDO human cDNA-sequencing project [27]. Blasting the FLJ12671 sequence against the nr database also identified hits on chromosomes 1 and 11, suggesting that this gene has been duplicated on several chromosomes. Ubiquitin conjugating enzyme E2D2 (UBE2D2) is a protein that targets abnormal or short-lived proteins for degradation by the 26S proteasome [28]. The EST sequences identified for the UBE2D2 gene on chromosome 6 are 93% identical to the human UBC4/5 gene located on chromosome 5 suggesting that this gene could be a duplicate as well. Adapter-related protein complex 3 (AP3) may be involved in intracellular protein transport [29]. Cytokeratin 8 is 95% identical to both the genomic and mRNA sequences of keratin. Vesicular membrane protein (P24), a previously characterized but unmapped gene [30] has been localized in intracellular organelles of highly differentiated neural cells and may have a role in the neural organelle transport system. ASSP2 is one of 12 pseudogenes of argininosuccinate synthetase encoded on 10
chromosomes with the only functional sequence residing on chromosome 9 [31]. Heterogeneous nuclear ribonucleoprotein A1 (HNRPA1) is also a pseudogene with three other copies on chromosomes 3, 13, and 20 [32]. The copy on chromosome 12q13.1 is thought to encode the gene responsible for the functional HNRPA1, which serves as a carrier for RNA during export to the cytoplasm [33].

The results of the qualitative RT-PCR, though not quantitative, were useful for characterizing the pattern of tissue expression (Figure 2). While most genes were about equally represented in the mRNAs of the twenty tissues in the panel, three genes, P24, NAD (+)-dependent succinic semialdehyde dehydrogenase (SSADH), and KIAA0319 had exceptional patterns. P24 was almost exclusively expressed in brain by RT-PCR, correlating with the origin of all 29 publicly accessible ESTs from brain cDNA libraries (Figure 2). RT-PCR suggested that the expression of SSADH was greatest in brain, though it was ubiquitously expressed in all tissues tested. Correspondingly, only 5 of 38 SSADH ESTs accessible on public domain servers were from brain cDNA libraries, with the remainder of mixed origin. KIAA0319 had strong signal from brain and cerebellum mRNA, reflected by the 10 of 29 publicly accessible ESTs originating from brain cDNA libraries.

**Discussion**

The primary goal of this study was to identify candidate genes surrounding the peak of transmission disequilibrium for RD on chromosome 6p and to characterize their patterns of expression. A secondary goal was to investigate the usefulness of the *in silico* approaches and specifically the dbEST database to identify and map new genes.

We identified 19 genes within 2 Mb of the peak of transmission disequilibrium – but are some better candidates for RD than others? The patterns of tissue expression as
profiled by RT-PCR and the frequency with which ESTs originated from brain cDNA libraries (Figure 2), serve to highlight five genes that are highly expressed in the brain: P24, SSADH, GPLD1, KIAA0386, and KIAA0319. Of these five, only one gene, SSADH, has been associated with a brain related phenotype. Two frameshift mutations, a G-to-T transversion in the intron 9 splice donor site, and a G-to-A transition in the intron 5 splice donor site, cause an exon to be skipped resulting in abnormal metabolism of GABA, an important neurotransmitter in the brain. The handful of described cases were originally diagnosed by anomalous GABA metabolites in the urine associated with developmental and speech delays, hyporeflexia, and behavioral problems including mild autism with clinical variation between affected family members [34]. There is no data, however, that links GABA or GABA metabolism to specific defects of reading independent of IQ. None of the other genes highly expressed in brain have associated diseases or clinical phenotypes. GPLD1 selectively hydrolyzes inositol phosphate linkages in vitro, releasing the protein bound to the plasma membrane via a glycosyl-phosphatidylinositol anchor into the cytosol [35]. P24 is a neuron specific membrane protein localized in intracellular organelles of highly differentiated neural cells and is involved in neural organelle transport. KIAA0386 encodes a protein that stimulates the formation of a non-mitotic multinucleate syncytium from proliferative cytotrophoblasts during trophoblast differentiation [36]. KIAA0319 encodes a protein of unknown function [37]. While HT012, an uncharacterized hypothalamus protein, could also be considered as a possible candidate gene, RT-PCR and EST searches (1 of 18 from brain cDNA libraries) do not suggest a high level or selective expression in the brain.

The in silico studies also identified candidate genes for other diseases that map to this region. The five brain candidate genes described above for RD are also reasonable candidates for the neurobehavioral disorders schizophrenia and ADHD. HSS results in the complete loss of scalp hair in childhood. Betz et al [21] described evidence for linkage to HSS with markers spanning D6S276 (400 Kb telomeric of JA04) through D6S1607 (5.6Mb centromeric of JA04). Neither the RT-PCR results nor the tissue origin for any single gene suggests any best candidates among the 19. Behçet’s disease is an autoimmune disorder characterized by a systemic vasculitis that affects the joints, all sizes and types of blood vessels, the lungs, the central nervous system, and the gastrointestinal tract [38]. There is evidence for linkage with markers spanning

Table 1: Genes encoded within 2 Mb of JA04 in order from telomere (top) to centromere.

| Gene         | Gene Size (kb) | mRNA Accession # | # of Exons | Mapping Source | Function                                                                 |
|--------------|----------------|------------------|------------|----------------|--------------------------------------------------------------------------|
| FLJ12671     | 15.885         | NM_030980        | 3          | This paper     | Function unknown                                                         |
| HNRPA1 (pseudogene) | 1.813      | XM_031157        | 1          | This paper     | Functions as carriers for RNA during export to the cytoplasm             |
| P24          | 21.142         | AF18980          | 5          | This paper     | Transmembrane protein; may have a role in vesicular trafficking          |
| RU2AS        | 79.995         | AF181722         | 7          | [46]           | Cellular defense response                                                |
| MR52L        | 20.629         | XM_041317        | 11         | NT_017021      | Transporter protein                                                       |
| GPLD1        | 61.339         | XM_041310        | 25         | [47]           | Hydrolizes inositol-PO4 linkage in proteins anchored by phosphatidylinositol glycerans |
| SSADH        | 39.639         | NM_001080        | 10         | [48]           | Required for 4-aminobutyric acid degradation                             |
| KIAA0319     | 102.051        | XM_011367        | 21         | NT_017021      | Function unknown                                                         |
| Cytokeratin 8| 2.206          | XM_069402        | 2          | This paper     | Regulation of exocrine pancreas homeostasis                              |
| TRAF         | 16.902         | AF223469         | 7          | [49]           | Inhibits NF-KappaB activation                                             |
| HT012        | 34.662         | AF220186         | 3          | NT_017021      | Uncharacterized hypothalamus protein                                    |
| FLJ12619     | 14.192         | XM_041290        | 5          | NT_017021      | Function unknown                                                         |
| AF3          | 1.667          | BC000804         | 3          | This paper     | May be involved in intracellular protein transport                       |
| Gemini       | 11.086         | XM_015895        | 9          | NT_017021      | Inhibits DNA replication during cell cycle S, G2, and M phase           |
| KIAA0386     | 106.682        | AB002384         | 23         | NT_017021      | Stimulates the formation of a non-specific multinucleate syncytium from proliferative cytotrophoblasts during trophoblast differentiation |
| Cyclophilin A| 0.81           | XM_094843        | 2          | NT_017021      | May be involved in protein folding and/or intracellular protein transport |
| ASSP2 (pseudogene) | 1.954    | XM_000845        | 1          | This paper     | Catalyzes the ATP-dependent ligation of citrulline to aspartate, forms argininosuccinate in the urea cycle |
| CMAR        | 56.489         | XM_041278        | 15         | [50]           | Converts the nucleotide sugar donor N-acetylnorleucaminic acid to N-glycolylleucaminic acid |
| NUP50       | 1.737          | XM_018531        | 1          | NT_017021      | Allows bi-directional transport of macromolecules through the nuclear envelope |
| UBE2D2      | 1.732          | XM_030770        | 1          | This paper     | Modifies proteins with ubiquitin for degradation                         |
| RPS10       | 1.428          | XM_069039        | 2          | This paper     | Ribosomal protein                                                        |

* The gene has been mapped on NCBI accession NT_017021 but no citation is present
nearly 28 Mb of 6p with JA04 in the middle [18]. Candidate genes for Behçet’s would include those expressed in lymphocytes or perhaps bone marrow such as AP3 and FLJ12671 and other immune related genes such as TRAF and RU2AS. These genes may also serve as candidates for other autoimmune disorders that map to this region such as IBD3.

Overall the in silico method for identifying genes in a specific genomic region worked well here, yielding a reasonable gene density of one per 95 kb. This method is highly dependent upon the quality of the information in dbEST. Any contamination from non-coding DNA, bacterial DNA, cDNA from other species, vectors or mitochondria DNA could generate false gene assignments. Fortunately, the high redundancy of EST hits in dbEST increased our confidence that any identification was likely physiologic and that the searches were sensitive. It is possible however, that our in silico approach may have missed some genes, in particular those with small ORFs and/or large 5-prime and/or 3-prime UTRs [39]. As dbEST expands over the next few years, new genes may be identified with repeated in silico searches, or with biophysical approaches such as cDNA hybridization [40], exon trapping [41] and amplification [42], or by identification of evolutionary conserved sequences [43] and HTF islands [44].

**Conclusion**

In summary, we examined 2 Mb surrounding the transmission disequilibrium peak with RD at short tandem repeat marker JA04 on chromosome 6p. In silico searches of the dbEST database identified 19 possible candidate genes. While tissue expression patterns suggest five candidates that are highly expressed in brain – one with a known association with neurological disease – neither the RT-PCR data nor the EST information can absolutely rule out any of the 19 as culpable candidates. We conclude therefore that in silico cloning is a powerful and effective technique for quickly identifying existing and novel genes, which can then be used to develop cDNA single nucleotide polymorphism markers (cSNPs) for pinpointing a more precise location of the 6p RD gene, and other disease genes that map to this region.

**Methods**

**In silico cloning**

The two million base pairs of index genomic sequence surrounding marker JA04 was downloaded from the NCBI website (accession number NT_017021). A perl script was written to parse the sequence into 200 files each containing 10 Kb segments in FASTA format. Repeat sequences were masked with the RepeatMasker program (RepeatMasker at http://ftp.genome.washington.edu/RM/RepeatMasker.html). Each masked file was sent to the blastc13 server as a query for searching the dbEST database using the BLAST algorithm [45]. Only ESTs with an identity of 93% or greater were considered; all other hits were discarded. Surviving ESTs were then used to search the nr database to identify parental cDNA or mRNA matches, which assembled 623 non-overlapping ESTs into 21 gene or pseudogene antecedents. The BLAST search of nr also showed hits against chromosome 6 BAC or PAC sequence. ESTs not mapping to chromosome 6 were also discarded. The final assemblies of EST and cDNA or mRNA sequences were aligned to the index NT_017021 genomic sequence using the Martinez Needleman-Wunsch algorithm in MegAlign (DNA Star, Lasergene) permitting identification of exon-intron boundaries, new exons (P24), and the direction of transcription relative to the telomere.

**Qualitative PCR**

A panel containing human total RNA from 20 different tissues was purchased from Clontech (BD Biosciences Clontech). RT-PCR was performed using the RETROscript first strand synthesis kit for RT-PCR (Ambion). 2µg of RNA was denatured with 0.5µM of random decamer
primers in a total volume of 12 µl, and heated to 77°C for three minutes and placed on ice. First strand cDNA was then synthesized with the addition of 2 µl 10 × RT buffer, 4 µl dNTP mix, 1 Unit RNase inhibitor, and 100 Units MMLV-RT in a total volume of 20 µl. The reaction was incubated for one hour at 44°C, and then 92°C for ten minutes and then stored at -20°C.

To check the quality of the first strand cDNA, 1 µl of cDNA was used in a PCR reaction to amplify the r1g/S15 ribosomal gene. 1.5 µl of 10 × PCR buffer (Qiagen), 250 µM of each dNTP, 0.5 Units of HotstarTaq polymerase (Qiagen) and 0.5 µM of primer (Ambion) were used in a 15 µl reaction. The reaction was denatured for 15 minutes at 95°C, then ten cycles of 94°C for 30 seconds, 65°C for 30 seconds (-1°C/cycle), and 72°C for 30 seconds, 20 additional cycles of 94°C for 30 seconds, 55°C for 30 seconds and 72°C for 30 seconds, and a final extension of 72°C for ten minutes. PCR reactions were performed in a MJ Research thermocycler. PCR products were electrophoresed on 2% agarose gels and stained with ethidium bromide. Primers (primer sequences are listed in Table 2) were designed to amplify the mRNA sequences of each of the 20 genes identified in the transcript map (Figure 1). Each amplicon was designed to be between 90 and 150 base pairs in length. For each amplicon, 1 µl of cDNA, 1.5 µl of 10 × PCR buffer (Qiagen), 250 µM of each dNTP, 0.5 units of HotstarTaq polymerase (Qiagen) and 0.5 µM of primer (Life Technologies) were used in a 15 µl reaction. PCR reactions were performed as above. One lane (Figure 2, lane 16) contained water during the RT-PCR step to check for RNA contamination. Products were electrophoresed on 2% agarose gels and stained with ethidium bromide.

Authors’ Contributions
ERL designed and performed the experiments and wrote the manuscript. HM aided in analyzing the data. JRG designed the experiments, analyzed the data, and edited the manuscript. All of the authors have reviewed the manuscript.

Note Added In Proof
Since submission of the manuscript for review, contig NT_017021 has been incorporated into NCBI contig NT_007592.13.

Acknowledgments
Support for JRG is through The Charles H. Hood Foundation, and National Institutes of Health (NIH) grants R01 DA12849, R01 DA12690, R01 NS43530, and U01 HD98002. Support for HM is through The Charles H. Hood Foundation. Support for EL is through The Robert Leet and Clara Guthrie Patterson Trust. The authors also wish to thank Thomas O’Reilly-Pol for help with characterizing NT_007592.13.

References
1. Rabin M, Wen XL, Hepburn M, Lubs HA, Feldman E and Duara R: Suggestive linkage of developmental dyslexia to chromosome 1p34-p36 Lancet 1993, 342:178.
2. Grigorenko EL, Wood FB, Meyer MS, Pauls JE, Hart LA and Pauls DL: Linkage studies suggest a possible locus for developmental dyslexia on chromosome 1p Am J Med Genet 2001, 105:120-129.

Table 2: Primer sequences for RT-PCR amplification of regional genes.

| Gene   | Forward primer          | Reverse primer          | Product size |
|--------|-------------------------|-------------------------|--------------|
| AP3    | GTTGTTGTGAGATCGACCAGAG | ATCATCAGGGGAATAGCTTCCA | 97 bp        |
| CMAH   | TGGAAAAACCAACACTTCTCA  | CAAATGAAAGACCATGAGA    | 105 bp       |
| Cyclophilin A | TGGAAAACAAAGCATAGCTG | CGAAAACCAACACTTCTCA | 105 bp       |
| Cytokeratin 8 | CCACTGAAGGAGGATCTGAT | TAATGCGGAGAAGAACTGAG | 95 bp        |
| ELF2D1 | ATGGGTTGTAAGGATTTGGA  | TGGGATGAAAGGACTTCCA    | 97 bp        |
| FLJ12619 | CTTTACGGGTGTTTTCAGG   | GAGGAATGAGGAACTTCTCA  | 100 bp       |
| FLJ12671 | CTCACCCAGCTCAGTTTTCT  | GCACATATCGAAGACCTTCCA | 100 bp       |
| Geminin | CACCTTTGATGAAAGACTG    | TCACTTTGATGAAAGACTG    | 100 bp       |
| GPD1   | CTTTACCTTGTTTTCAGG    | CTCACCTTTGATGAAAGACTG | 100 bp       |
| HT012  | TGGAAAACAAAGCATAGCTG  | CGAAAACCAACACTTCTCA   | 100 bp       |
| KIAA0319 | GGTACATACCATACGTTTCT | CCCACCTTTGATGAAAGACTG | 100 bp       |
| KIAA0386 | TATGACGGTGAAGGAGCTCAT | AATTAGGCGGACTTCTTAAG  | 117 bp       |
| RPS10  | CATGACGGTGAAGGAGCTCAT | AAATGAGGGGACTTACTTTGA | 108 bp       |
| MRS2   | ATGAGGCAAGCAGAAGGACAC | GATGTGACGGCTTTGTAGGA  | 109 bp       |
| NUP50  | GAGAGTGGAGGATGTCGCGAT | GTGAGGAGGACTTACTTTGC | 117 bp       |
| P24    | TCTCAGGGTTGTAAGGATTTG | GAGGAATGAGGAACTTCTCA  | 100 bp       |
| RU2AS  | TGGAAAACAAAGCAGCCATG  | TCACTTTGATGAAAGACTG    | 105 bp       |
| SSADH  | TGGAAAACAAAGCAGCCATG  | TCACTTTGATGAAAGACTG    | 105 bp       |
| TRAF   | TGGAAAACAAAGCAGCCATG  | TCACTTTGATGAAAGACTG    | 105 bp       |
| S15    | TGGAAAACAAAGCAGCCATG  | TCACTTTGATGAAAGACTG    | 105 bp       |
40. Parimoo S, Patanjali SR, Shukla H, Chaplin DD and Weissman SM: BMC Genomics
41. Duyk GM, Kim SW, Myers RM and Cox DR:
42. Buckler AJ, Chang DD, Graw SL, Brook JD, Haber DA, Sharp PA and
50. Irie A and Suzuki A:
49. Py I, Declercq W, Ibrahim A, Michiels C, Van Rietshoten JG, Dewulf N, de Boer M, Vandenabeele P, Huylenbroeck D and Remacle JE: TTRAP, a novel protein that associates with CD40, tumor necrosis factor (TNF) receptor-75 and TNF receptor-associated factors (TRAFs), and that inhibits nuclear factor-kappa B activation J Biol Chem 2000, 275:18586-18593.
50. Irie A and Suzuki A: CMP-N-Acetylneuraminic acid hydroxylase is exclusively inactive in humans Biochim Biophys Res Commun 1998, 248:330-333.