From DEPARTMENT OF BIOSCIENCES AND NUTRITION
Karolinska Institutet, Stockholm, Sweden

MOLECULAR STUDIES OF DYSLEXIA: REGULATION AND FUNCTION OF DYX1C1

Kristiina Tammimies

Karolinska Institutet

Stockholm 2011

Published by Karolinska Institutet.

ISBN 978-91-7457-511-8

Printed by
REPROPRINT AB
Stockholm 2011
www.reproprint.se
Gårdsvägen 4, 169 70 Solna

One never notices what has been done;
one can only see what remains to be done.
(Marie Curie, letter to her brother, 1894)

To all my loved ones

ABSTRACT

Developmental dyslexia is a specific reading disability characterized by unexpected difficulty in reading and writing despite adequate intelligence, education, normal senses and social environment. It is the most common childhood learning disorder affecting five to ten percent of school age children and it is more common among boys than girls. The core deficit in dyslexia is believed to involve phonological processing, the lowest level of the language system needed for reading. Dyslexia has a neurological basis demonstrated by anatomical and functional brain studies, in which differences have been found in the brains of dyslexic readers compared to normal readers. Subtle disturbances in neuronal migration during early brain development have been suggested to be one of the mechanisms leading to dyslexia.

Dyslexia has a complex genetic basis that has been investigated by extensive family, twin- and molecular genetic studies. To date, many chromosomal loci, including the nine official dyslexia loci, have been linked to dyslexia, and a number of susceptibility genes within those regions have been identified. At least four of these candidate genes are involved in neuronal migration and brain development, otherwise their function it not well understood.

The aim of this thesis was to study the regulation and function of the first dyslexia susceptibility gene *DYX1C1*. The *DYX1C1* gene was identified when it was disrupted by a translocation segregating with dyslexia in one family. Since then, many association studies have supported its role in the etiology of dyslexia and general reading ability. In rodents, embryonic knockdown of *Dyx1c1* results in deficits in neuronal migration leading to ectopias in the neocortex and hippocampus, and impairments in performing tasks related to learning and memory.

In Paper 1, we characterized three dyslexia associated single nucleotide polymorphims in the regulatory regions of *DYX1C1* and identified regulatory proteins binding to the genomic region upstream of the translation start site. We showed that these changes could have functional consequences and therefore could explain the association signal. In Papers II and III, we connected DYX1C1, both function and its regulation, to estrogen signaling. The expression of DYX1C1 increased after treatment with the steroid hormone, 17β-estradiol, which was due to the regulatory effect of the estrogen receptor β and TFII-I (III). Furthermore, we demonstrated that DYX1C1 interacts with the estrogen receptors α and β with functional consequences (II). In Paper IV, we scrutinized the function of DYX1C1 by characterizing the global gene-expression patterns after manipulating its expression levels in a neuroblastoma cell line and by identifying its protein interactions partners. By this means, we connected DYX1C1 to molecular pathways relevant to neuronal migration and nervous system development. For instance, the expression of neuronal migration genes RELN and DCX was changes after manipulating DYX1C1 levels. In addition, we studied the random cell migration of neuroblastoma cells after perturbation of DYX1C1 levels to confirm that the identified pathways and connections are functional. Indeed, DYX1C1 affects the velocity of the random cell migration and the protein domains in the C-terminus of DYX1C1 are needed for this.

From the findings in this thesis, we can conclude that DYX1C1 is involved in several interesting molecular pathways and we provide starting points for future studies. In addition, we strengthen and further develop some of the already existing theories of the biological causes of dyslexia.

LIST OF PUBLICATIONS

I. Isabel Tapia-Páez, **Kristiina Tammimies,** Satu Massinen, Ananda L. Roy and Juha Kere.
The complex of TFII-I, PARP1, and SFPQ proteins regulates the *DYX1C1* gene implicated in neuronal migration and dyslexia.
FASEB journal, 2008, Aug;22(8):3001-9

II. Satu Massinen*, **Kristiina Tammimies***, Isabel Tapia-Páez, Hans Matsson, Marie-Estelle Hokkanen, Ola Söderberg, Ulf Landegren, Eero Castrén, Jan-åke Gustafsson, Eckardt Treuter and Juha Kere.
Functional interaction of DYX1C1 with estrogen receptors suggests involvement of hormonal pathways in dyslexia.
Human Molecular Genetics, 2009 Aug 1;18(15):2802-12
* Authors contributed equally

III. **Kristiina Tammimies**, Isabel Tapia-Páez, Gustaf Rosin, Juha Kere, Jan-Åke Gustafsson and Ivan Nalvarte.
Estrogen receptor beta enhances the transcription of dyslexia candidate gene *DYX1C1* in a 17beta-estradiol dependent manner.
Manuscript

IV. **Kristiina Tammimies,** Morana Vitezic, Hans Matsson, Sylvie Le Guyader, Thomas Burglin, Tiina Öhman, Staffan Strömblad, Carsten O. Daub, Tuula A. Nyman, Juha Kere and Isabel Tapia-Páez.
The Molecular Networks of the Dyslexia Gene *DYX1C1* Show Connection to Neuronal Migration Genes and Cytoskeletal Proteins
Manuscript

ADDITIONAL PUBLICATIONS

Matsson H, **Tammimies K**, Zucchelli M, Anthoni H, Onkamo P, Nopola-Hemmi J, Lyytinen H, Leppanen PH, Neuhoff N, Warnke A, Schulte-Körne G, Schumacher J, Nöthen MM, Kere J, Peyrard-Janvid M.
SNP variations in the 7q33 region containing *DGKI* are associated with dyslexia in the Finnish and German populations.
Behav Genet. 2011 Jan; 41(1):134-40.

Massinen S, Hokkanen ME, Matsson H, **Tammimies K**, Tapia-Páez I, Dahlström-Heuser V, Kuja-Panula J, Burghoorn J, Jeppsson KE, Swoboda P, Peyrard-Janvid M, Toftgård R, Castrén E, Kere J.
Increased Expression of the Dyslexia Candidate Gene *DCDC2* Affects Length and Signaling of Primary Cilia in Neurons.
PLoS One. 2011; 6(6):e20580. Epub 2011 Jun 16.

CONTENTS

1 Popular scientific summary in Swedish and Finnish 1
1.1 Populärvetenskaplig sammanfattning 1
1.2 Populaaritieteellinen tiivistelmä 3
2 Introduction 5
2.1 Reading 5
2.2 Dyslexia 6
2.2.1 History and definition 6
2.2.2 Symptoms and diagnosis of dyslexia 7
2.2.3 Prevalence 8
2.2.4 Etiology 8
2.2.5 Neurobiology 9
2.3 Neuronal migration 12
2.3.1 Normal neuronal migration 12
2.3.2 Neuronal migration in disease 13
2.4 The human genome 14
2.4.1 Sequence variations in the genome 15
2.4.2 The central dogma of molecular biology 16
2.4.3 Beyond the central dogma 18
2.5 Gene discovery in complex disorders 19
2.5.1 Linkage and association analysis 19
2.5.2 Other approaches 20
2.5.3 Genome-wide association studies 20
2.5.4 The future of gene mapping in complex disorders 21
2.6 Candidate loci and genes for dyslexia 22
2.6.1 DYX1C1 24
2.6.2 ROBO1 29
2.6.3 DCDC2 30
2.6.4 KIAA0319 31
2.6.5 Additional genes 32
2.7 Estrogen receptors and their signaling 34
2.7.1 Estrogen signaling in brain 35
3 Aim of the thesis 36
4 Methodology 37
4.1 Cell lines 37
4.2 Gene expression analysis 38
4.2.1 Quantitative real-time polymerase chain reaction (I,III,IV) 38
4.2.2 Expression microarrays (IV) 38
4.3 Protein analysis 39
4.3.1 Western blot (I, II, III and IV) 39
4.3.2 Electrophoretic mobility shift assay (I) 39
4.3.3 Immunocytochemistry (I, II, III and IV) 39
4.3.4 In situ proximity ligation (II, IV) 40
4.3.5 Immunoprecipitation (II, III and IV) 40
4.3.6 Protein identification by LC-MS/MS (I, IV) 41
4.4 Live-cell imaging 41

4.5 Luciferase assays 42
4.6 Bioinformatic tools 42
5 Results and discussion 43
5.1 The regulation of *DYX1C1* expression (I and III) 43
5.1.1 Transcription factors binding to the *DYX1C1* regulatory region 45
5.2 DYX1C1 and estrogen receptor signaling (II and III) 47
5.3 The effect of DYX1C1 to global gene expression (IV) 49
5.4 Evolutionary analysis of DYX1C1 protein: Identification of a new protein domain 52
5.5 Protein interacting partners for DYX1C1 (IV) 53
5.6 Localization of the DYX1C1 protein (I, II, III and IV) 55
5.7 The role of DYX1C1 in cell migration 56
6 Concluding remarks and future perspectives 59
7 Acknowledgements 61
8 References 64

LIST OF ABBREVIATIONS

CCC	Cilia-centrosome complex
CD/CV	Common diseases/common variant
ChIP	Chromatin immunoprecipitation
CNV	Copy-number variation
Co-IP	Protein complex immunoprecipitation
CP	Cortex plate
CR	Cajal-Retzius
DNA	Deoxyribonucleic acid
EMSA	Electrophoretic mobility shift assay
ER	Estrogen receptor
ERE	Estrogen responsive element
GWAS	Genome-wide association study
GO	Gene ontology
HGP	Human genome project
HRP	Horseradish peroxidase
IBD	Identical by descent
ICC	Immunocytochemistry
iPSC	Induced pluripotent stem cell
isPLA	*In situ* proximity ligation assay
LC-MS	Liquid chromatography-mass spectrometry
LD	Linkage disequilibrium
LGE	Lateral ganglionic eminence
MGE	Medial ganglionic eminence
MRI	Magnetic resonance imaging
NMD	Neuronal migration disorder
qRT-PCR	Quantitative real-time polymerase chain reaction
RNA	Ribonucleic acid
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SLI	Specific language impairment
SNP	Single nucleotide polymorphism
SRE	Serum response element
SSD	Speech sound disorder
SSR	Simple sequence repeat
SV	Structural variant
SVZ	Subventricular zone
TF	Transcription factor
TFBS	Transcription factor binding site
TPR	Tetratricopeptide repeat
TRE	Transcription regulatory element
UTR	Untranslated region
VZ	Ventricular zone
WBS	Williams-Beuren syndrome

1 POPULAR SCIENTIFIC SUMMARY IN SWEDISH AND FINNISH

1.1 POPULÄRVETENSKAPLIG SAMMANFATTNING

Att lära sig läsa och skriva är den första utmaningen ett barn möter i skolan. För att uppnå detta måste barnen ha förmågan att kunna avkoda text till språk, vilket kräver snabba och automatiserade processer i hjärnan. För ungefär en tiondel av barnen uppstår dock svårigheter att lära sig att läsa och skriva på grund av dyslexi. Dyslexi betyder ”svårigheter med ord” och är ett mångfasetterat problem som drabbar fler pojkar än flickor och uppstår trots normal intelligens, utbildning eller social miljö. Till följd av att större delen av all kunskap överförs genom skrift, i form av tidningar, böcker och Internet, är läs- och skrivförmåga mycket viktigt i dagens samhälle. Om ett dyslektiskt barn inte får rätt diagnos och adekvata pedagogiska åtgärder kan detta leda till sämre utbildningsmöjligheter och på sikt att han eller hon blir utestängd från samhället.

Dyslexi har studerats ur olika infallsvinklar under många decennier och idag vet vi att åkomman beror på en ärftlig inlärningsstörning med neurobiologisk bakgrund. I studier där hjärnor hos individer med dyslexi jämförts med individer med normal läsförmåga har man observerat skillnader i både hjärnans struktur och aktivitet, och där nedsatt aktivitet ses i vänstra hjärnhalvan hos dyslektiker. Detta är intressant på grund av att det är den del av hjärnan där det talade och skrivna språket behandlas. När mar har analyserat hjärnon hos avlidna personen med dyslexi har man även hittat avvikelser där grupper av nervceller har gått vilse och hamnat fel i hjärnbarken. Detta beror troligen på att dessa nervceller vandrar fel under ett tidigt utvecklingsstadium av hjärnan.

Det har under lång tid varit känt att dyslexi är vanligare i vissa familjer än andra. Med hjälp av familje- och tvillingstudier har man kunnat estimera att klart mer än hälften av dyslexins uppkomst kan förklaras med gener. Resten kan man förklara med olika miljöfaktorer- vilka dessa är vet man dock mycket litet om i nuläget. Som tur är vet vi mycket mer om den genetiska bakgrunden. Vår arvsmassa (genomet) innehåller genetiska markörer (vissa variabla DNA-regioner) vars nedärvning mellan generationer kan spåras med kopplingsanalys. Om någon av dessa markörer finns hos personer med dyslexi men inte hos de familjemedlemmar som läser normalt kan antas att det omgivande området i DNA ha ärvts från familjemedlemmar med dyslexi. Denna DNA-region innehåller därför en predisponerande genetisk defekt. Genom sådana genetiska studier (kopplingsanalys) har åtminstone nio kromosomregioner förknippats med dyslexi, och i dessa regioner har dussintals riskgener hittats. Av dessa har DYX1C1, DCDC2 och KIAA0319 verifierats till riskgener i flera studier. Dessa tre gener tillsammans med ROBO1 har också bidragit till att förstå den neurobiologiska bakgrunden till dyslexi. Med hjälp av djurstudier har man kunnat visa att DYX1C1, DCDC2 och KIAA0319 -generna är nödvändiga för utvecklingen av en normal hjärna hos råttor och utan dessa gener hämmas den neuronala migreringen till hjärnbarken. ROBO1 är vidare involverad i hjärnans cellvandring och är även en mycket viktig receptor-gen, som deltar i bildandet av hjärnbalken som förbinder hjärnhalvorna.

Syftet med min avhandling var att utforska regleringen och funktionen av DYX1C1-genen för att på så sätt försöka förstå dess roll i utvecklandet av dyslexi. Vi visste sen tidigare att DYX1C1-genen deltar i hjärnans utveckling, men dess exakta roll och biokemiska signalvägar har inte studerats tidigare. Ett sätt att studera gener är använda mänskliga celler som odlas på laboratoriet. I varje cell i vår kropp har vi 22

kromosompar och två könskromosomer, som innehåller cirka 22 000 proteinkodande gener. Vanligtvis uttycks bara del av dessa gener i en cell, vilket betyder att genernas uttryck varierar mellan olika typer av celler. Uttrycket av gener i celler regleras mycket noga med regleringsproteiner som binder till specifika DNA-regioner. Om dessa regleringsregioner innehåller fel eller förändringar kan det bidra till att uttrycket av genen förändras. Ändringar i genuttrycket kan var orskaken till ökad risk för vissa sjukdomar.

I studie I har vi kunnat visa att det finns minst tre olika förändringar i reglerande regioner av DYX1C1 som är kopplade till en ökad risk för dyslexi och att dessa förändringar påverkar bindningen av proteiner till den reglerande regionen och därmed uttrycket av genen. Baserat på dessa resultat kan vi anta att förändring i uttrycket av DYX1C1 kan leda till ökad risk av dyslexi.

Genom att utforska DYX1C1-genens aktivitet har vi i studie II och III kunnat visa att det finns en koppling mellan DYX1C1-genen och östrogenreceptorer. Östrogen och dess receptorer har många funktioner; till exempel att modifiera minne och kognitiva egenskaper i hjärnan, inte bara hos kvinnor utan också hos män. Vi kunde visa att när man exponerar celler för östrogen leder det till en snabb ökning av DYX1C1-uttrycket, som tur leder till att DYX1C1-proteinet binder till östrogenreceptorer och reglerar deras verksamhet. Denna koppling är mycket intressant med tanke på att dyslexi är vanligare hos pojkar än hos flickor.

Molekyler samverkar med varandra för att bilda olika nätverk vars uppgift är att kontrollera och utföra funktioner i celler och vävnad. I studie IV har vi använt olika metoder för att ta reda på vilka gener och genprodukter som är kopplade till DYX1C1. När vi byggde ett nätverk av gener och protein runt DYX1C1 kunde vi koppla DYX1C1 till gener som är viktiga för hjärnans och nervsystemets utveckling. DYX1C1-proteinet binder också till flera andra proteiner kopplade till cellstruktur, underhåll och transport. Med dessa resultat kunde vi kombinera funktionen av DYX1C1 med molekylära signalvägar som är viktiga i hjärnans utveckling, och med detta stärka hypotesen att milda störningar i vandringen av neuroner kan vara en del av den neurobiologiska orsaken bakom dyslexi.

Eftersom dyslexi orsakas av många gener tillsammans med miljöfaktorer är det mycket intressant att studera de redan identiferade riskgener och deras effekter på biokemiska signalvägar. Genom att studera genernas effekt på dessa signalvägar kan man också hitta fler riskgener. För den allmänna forskningen är det också viktigt att hitta gener bakom dyslexi. Detta för att kunna få mer information om hjärnans normala utveckling liksom utvecklingen av gener som gör det möjligt för människor att läsa. På sikt skulle förståelse av den biologiska bakgrunden till dyslexi bidra till att utveckla metoder för att kunna ställa en tidigare diagnos och med detta kunna hjälpa barn med dyslexi att få den hjälp de behöver så tidigt som möjligt. Detta är viktigt då tidig diagnos och specialundervisning hjälper dyslektiska barn att undvika de svårigheter som uppkommer i skolan och på arbetet till följd av dyslexi.

1.2 POPULAARITIETEELLINEN TIIVISTELMÄ

Luku- ja kirjoitustaidon oppiminen on ensimmäisen kouluvuoden tärkeimpiä saavutuksia. Lukeminen on monimutkainen prosessi, jossa aivot tekevät nopeaa työtä kirjoitetun tekstin kääntämiseksi puhutuksi kieleksi. Nykypäivän yhteiskunnassa riittävä luku- ja kirjoitustaito on erittäin tärkeä, koska suurin osa tiedosta välitetään kirjoitetussa muodossa sanomalehtien, kirjojen ja internetin kautta. Lukeminen ei kuitenkaan ole helppoa kaikille, sillä lähes joka kymmenennellä lapsella on vaikeuksia oppia lukemaan lukihäiriön takia. Tämä monitasoinen ongelma on yleisempi pojilla kuin tytöillä, ja se ilmenee normaalista älykkyydestä, koulutuksesta tai sosiaalisesta ympäristöstä riippumatta. Lukihäiriö voi aiheuttaa lapselle sosiaalisia ongelmia ja hidastaa oppimista, jos sitä ei tunnisteta ajoissa ja järjestetä riittävää tuki- ja erityisopetusta.

Nykyisin tiedetään, että lukihäiriö on perinnöllinen neurologinen häiriö, ja sen taustoja on selvitetty jo monien vuosikymmenien ajan eri tutkimusalojen näkökulmasta. Lukihäiriöisten aivoissa on löydetty esimerkiksi eroja aktivointikaavoissa ja aivorakenteissa verrattuna normaalilukijoihin. Erityisesti lukihäiriöisillä on havaittu alentunutta aktiivisuutta vasemman aivopuoliskon alueilla, joissa puhuttu ja kirjoitettu kieli prosessoidaan. Lisäksi joidenkin lukihäiriöisten aivojen solurakenteissa on löydetty poikkeavuuksia aivokuoressa. Nämä poikkeavuudet johtuvat siitä, että hermosoluryhmiä on eksynyt väärälle paikalle aivokuoressa jo sikiövaiheessa.

Jo viime vuosisadan alussa tutkijat huomasivat, että lukihäiriö kasaantui tiettyihin sukuihin. Sen jälkeen lukuisat perhe- sekä kaksostutkimukset ovat tuoneet varmuuden sen perinnöllisestä taustasta. On myös voitu arvioida, että yli puolet lukihäiriön taustasta voidaan selittää geenien vaikutuksella. Loput voidaan selittää ympäristötekijöillä, vaikka lukihäiriölle altistavista ympäristötekijöistä ei juuri olekaan tietoa.

Lukihäiriön perinnöllisestä taustasta sen sijaan tiedetään enemmän. Perimässämme eli genomissa on ns. geneettisiä markkereita eli tiettyjä muuntelevia DNA-alueita, joiden periytymistä voidaan jäljittää kytkentäanalyysin avulla. Jos jokin tutkittavista markkereista esiintyy lukihäiriöisillä mutta ei perheenjäsenillä, jotka lukevat normaalisti, voidaan olettaa, että sitä ympäröivä DNA-alue on periytynyt suvussa lukihäiriöisillä. Tämä DNA-alue sisältää siis altistavan geenivirheen. Näin on pystytty yhdistämään lukihäiriöalttius moniin kromosomialueisiin, joista yhdeksän on saanut virallisen lukihäiriölokuksen nimen DYX1-9. Tähän mennessä näiltä alueilta on löydetty kymmeniä alttiusgeenejä, joista kolme, DYX1C1, DCDC2 sekä KIAA0319, on pystytty varmistamaan alttiusgeeneiksi useissa tutkimuksissa. Nämä kolme geeniä sekä lisäksi ROBO1 ovat auttaneet ymmärtämään dysleksian biologista taustaa. Eläinkokeiden avulla on todettu, että DYX1C1, DCDC2 ja KIAA0319 -geenit ovat tarpeen rotan normaalissa aivojen kehityksessä, ja että ilman niitä hermosolujen kulkeutuminen omalle paikalleen aivokuoreen estyy. ROBO1 osallistuu myös aivosolujen vaellukseen, mutta on myös erittäin tärkeä aksoninohjauksen reseptorigeeni, joka osallistuu aivopuoliskoja yhdistävän aivokurkiaisen muodostumiseen.

Tämän väitöskirjan tavoitteena oli selvittää DYX1C1-geenin säätelyä ja toimintaa solutasolla, mikä voisi auttaa ymmärtämään sen roolia lukihäiriön synnyssä. DYX1C1-geenin tiedetään siis olevan tarpeen aivojen kehityksen aikana, mutta sen tarkkaa tehtävää sekä biokemiallisia reittejä ei ole ennen tutkittu. Tämän väitöskirjan

tutkimuksiin käytettiin suurimmaksi osaksi solulinjoja, joita voidaan kasvattaa soluviljelyn avulla rajattomasti. Näin pystyttiin helposti tutkimaan ja säätelemään geenien ilmentymistä (ekspressiota) ilman eläinkokeita. Ihmisen jokaisessa solussa on 22 kromosomiparia ja kaksi sukupuolikromosomia, jotka sisältävät noin 22 000 proteiinia koodaavaa geeniä. Yleensä vain osaa näistä geeneistä luetaan eli niiden ilmentyminen vaihtelee eri solutyypeissä. Geenien ilmentymistä soluissa säädellään siis hyvin tarkasti erilaisten säätelyproteiinien avulla. Nämä proteiinit sitoutuvat tiettyihin DNA- kohtiin, ja jos näissä kohdissa esiintyy emäsvirheitä, voi geenin ilmentyminen muuttua.

Ensimmäisessä osajulkaisussa pystyttiin todistamaan, että näin on myös DYX1C1 kohdalla. Sen säätelyalueilla on ainakin kolme emäsmuutosta, jotka on yhdistetty aikaisemmin lukihäiriöalttiuteen, ja jotka vaikuttavat säätelyproteiinien sitoutumiseen ja näin myös geenin ilmentymiseen. Lisäksi monta aivoissa esiintyvää säätelyproteiinia, jotka sitoutuvat DYX1C1-geenin säätelyalueille tunnistettiin tämän väitöskirjan tutkimissa. Näiden tulosten perusteella voidaan olettaa, että hienoinenkin muutos DYX1C1-geenin toiminnan säätelyssä voisi johtaa lukihäiriöalttiuteen.

Tutkimalla DYX1C1-geenin toimintaa pystyimme yhdistämään sen estrogeenireseptoreihin sekä niiden vaikutusreittiin osajulkaisuissa II ja III. Solujen altistaminen estrogeenille aiheuttaa nopean kasvun DYX1C1 ilmentymisesssä, mikä johtaa siihen, että DYX1C1-proteiini sitoutuu estrogeenireseptoreihin ja säätelee näiden toimivuutta. Tämä yhteys on erittäin mielenkiintoinen siinä valossa, että lukihäiriö on yleisempää pojilla kuin tytöillä.

Molekyylit toimivat yhdessä muodostaen verkostoja, joiden tehtävänä on kontrolloida ja toteuttaa tehtäviä soluissa. Osajulkaisussa IV käytettiin menetelmiä, joilla voitiin selvittää, mitkä geenit ja niiden tuotteet eli proteiinit ovat yhteydessä DYX1C1-geenin kanssa. Rakentamalla geeni- ja proteiiniverkostoa DYX1C1:n ympärille pystyttiin yhdistämään se geeneihin, jotka ovat tärkeitä sekä aivojen että hermoston kehityksen kannalta, ja jotka säätelevät hermosolujen vaellusta (migraatiota). DYX1C1-proteiini sitoutuu useisiin solun rakenteen ylläpito- ja moottoriproteiineihin, joita tarvitaan myös solujen vaellukseen. Näin on pystytty yhdistämään aivokuoren kehityksen kannalta tärkeitä molekyylivaikutusreittejä DYX1C1-geenin toimintaan. Nämä tulokset vahvistavat hypoteesia siitä, että lukihäiriön neurobiologinen syy on osaksi häiriintynyt hermosolujen vaellus aikaisessa kehityksessä.

Koska lukihäiriö on monitekijäinen ominaisuus, on erittäin kiinnostavaa tutkia alttiusgeenejä ja niiden biokemiallisia vaikutusreittejä. Tutkimalla näitä vaikutusreittejä voidaan myös löytää lisää alttiusgeenejä. Yleisesti ottaen on myös kiinnostavaa löytää lukihäiriön takana olevia geenejä, ja näin saada enemmän tietoa aivojen normaalista kehittymisestä sekä siihen tarvittavista geeneistä. Näin voimme myös ymmärtää paremmin, mitkä geenit ovat osallistuneet ihmisaivojen monimutkaisten rakenteiden syntymiseen evoluution aikana ja näin mahdollistaneet erityisten taitojen, kuten lukemisen, oppimisen. Lukihäiriön biologisen taustan ymmärtäminen voi edistää myös sen varhaisen diagnosoinnin kehittämistä, mikä mahdollistaisi lukihäiriön aikaisen toteamisen ja tukitoimien aloittamisen ja auttaa välttämään näin lukihäiriön tuottamia vaikeuksia sekä koulussa että työelämässä.

2 INTRODUCTION

2.1 READING

In today's society, one of the most important achievements is to learn how to read and write. Without sufficient literacy skills, one may be left behind in the society and develop deep social and financial consequences, since almost all information is given in print not only in magazines, newspapers and information brochures but also on the internet.

Reading is a complex process containing both physiological and cognitive components and contrary to the spoken language it is acquired and needs to be taught. Learning to read requires multiple skills: First, one needs to recognize that spoken language can be segmented into smaller elements (phonemes); this process is also referred as the phonological awareness. Then, one has to learn how to identify letters and the corresponding sounds and how these can be printed into words (orthographic skills), followed by learning the vocabulary and extracting meaning from the printed words (semantic knowledge) (Shaywitz and Shaywitz 2008). Automatization of the mentioned skills and short-term memory are needed for fluent reading and writing (Price 2000).

Mature reading is performed by a left-hemisphere network of the brain consisting of the frontal, temporoparietal and occipitotemporal cortical systems (Shaywitz and Shaywitz 2008). Behavioral and brain imaging studies have shown that reading consists of multiple interacting subcomponents and different activation patterns depending on the reading task (Price 2000). A simple model of how the brain processes printed text in to speech was already made after initial observations by Dejerine when he studied different alexias (Geschwind 1965). The model shows that when we read, our brain first codes the printed word in the primary visual cortex, from where this information is passed on to the angular gyrus, located in the junction of the parietotemporal and occipitotemporal systems, to be decoded as semantic information. The well-organized angular gyrus passes on the information to Wernicke's and further to Broca's area to be interpreted as the corresponding auditory word (Figure 1) (Price 2000; Turkeltaub *et al.* 2003). Reading is an extremely rapid process, just think how many words you have already decoded while reading this chapter!

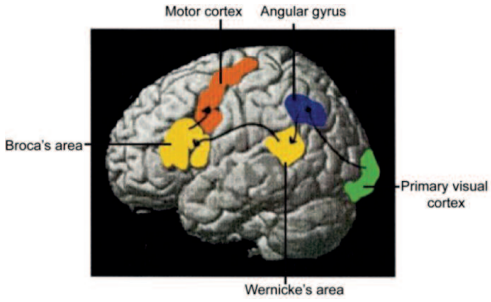

Figure 1 The flow of reading process in the brain. The primary input is coded in the visual cortex where it is then passed on to the angular gyrus, Wernicke's area and Broca's area. The final destination is in the motor cortex where the input is decoded back to spoken word. Modified from *J. Anat* (197) C.J Price, The anatomy of language: contributions from functional neuroimaging. Copyright 2000, with permission from John Wiley and Sons.

2.2 DYSLEXIA

Humans have used the printed word for over 5000 years (Lawler 2001), and there have probably always been people that struggle with decoding of the printed symbols back to spoken language, a disorder that we today call *dyslexia.* Dyslexia can be acquired (e.g. caused by a brain damage) or developmental. The disorder has fascinated thousands of researchers and clinicians in different fields such as neurophysiology, neuropathology, psychology, educational sciences and molecular genetics. They search for answers to questions like: What is dyslexia? What is the underlying cause and why it is specific to reading? How can we identify dyslexia before the child fails to learn to read? How can we teach dyslexic children to cope with the disorder? (Adapted from (Nicolson 2008)). These are all very important questions that need to be answered, and the whole picture can only be completed by translational research across these different research fields. This chapter briefly summarizes what is known about the history, etiology and neurobiology of dyslexia.

2.2.1 History and definition

The first notes of reading and writing difficulties as a medical condition are from the early 19th century, when physicians studied different aphasias such as motor aphasia (difficulty in reading) and alexia (difficulty in writing). In 1877, the term *"word blindness"* was taken into use by the German physician Adolph Kussmaul when he described a patient with severe reading difficulties despite normal intelligence. In addition, Morgan and Hinshelwood reported cases of word blindness occurring in children. These studies were based on cases where the condition was caused by neurological damage that resulted in loss of reading ability. Ten years later, the German ophthalmologist R. Berlin introduced the term dyslexia, which is derived from both Latin and Greek. The Latin origin is dys (dis=difficult) + legere (to read); or Latin dys + Greek lexis (speech). Thus, dyslexia would mean difficulty with reading and speaking. In 1923, Samuel Orton further defined the condition when he described the first theory of *specific learning disability*, in which he emphasized the fact that listening, speaking, reading and writing all together form a unity of language. (Richardson 1992; Grigorenko 2001)

To date the definition of dyslexia (also known as a specific reading disability) has been specified and sharpened. According to the International Dyslexia Association (IDA) and the National Institute of Health (NIH) the definition is (2002):

> "Dyslexia is a specific learning disability that is neurological in origin. It is characterized by difficulties with accurate and/or fluent word recognition and by poor spelling and decoding abilities. These difficulties typically result from a deficit in the phonological component of language that is often unexpected in relation to other cognitive abilities and the provision of effective classroom instruction. Secondary consequences may include problems in reading comprehension and reduced reading experience that can impede the growth of vocabulary and background knowledge."

2.2.2 Symptoms and diagnosis of dyslexia

The spectrum of problems, both cognitive and behavioral, in developmental dyslexia (referred here on only as dyslexia) is broad, spanning from spelling errors to severe difficulties in the reading of single words and reading fluency. Dyslexia is therefore a very complex phenotype. It has been discussed that dyslexics do not form a homogenous phenotypic population, but rather fall into a number of distinct subgroups (Castles and Coltheart 1993; King *et al.* 2007).

Characteristic symptoms in dyslexia are problems in segmenting words into phonemes, keeping linguistic material in short-term memory, reading and writing non-words and slow naming of letters or objects. The symptoms has been mostly explained by deficits in phonological awareness as noted also in the official definition (Ramus and Szenkovits 2008). Dyslexic children also show significant underachievement on tests of single-word reading and reading fluency. The double-deficit hypothesis of dyslexia takes into account this combination of phonological and processing speed deficits (Vukovic and Siegel 2006). In addition, there are many more theories in the literature that provide explanations of the underlying biological and cognitive causes of dyslexia. The two major ones are the cerebellar and magnocellular theories claiming that dysfunction in the cerebellum or in the magnocellular system could be the cause of dyslexia (Stein 2001; Stoodley and Stein 2011).

Dyslexia is usually diagnosed in primary school after the reading difficulties of the child have been noticed by the teachers and parents. The proper diagnosis is done with a battery of cognitive tests measuring reading and writing ability and various subcomponents of dyslexia (Skiba *et al.* 2011). Since it is believed that dyslexic children represent the lower end of the normal distribution of reading capabilities, there is no distinct cut-off between the dyslexic children and normal readers (Shaywitz *et al.* 1992). Therefore, the clinical diagnosis needs to be done by applying thresholds to the psychometric measures that are normally distributed in the general population. The common criterion is to have at least two standard deviation (SD) discrepancies between the observed reading ability and what is expected on the basis of age and intelligence quotient (IQ) (Fisher and DeFries 2002).

2.2.3 Prevalence

Dyslexia is the most common childhood learning disability affecting 5-10 % or even up to 17.8 % of the population (Shaywitz *et al.* 1992; Shaywitz 1998; Katusic *et al.* 2001). The prevalence estimates for dyslexia can vary depending on the language, the age and the use of different diagnostics criteria (Grigorenko 2001). It has been suggested that especially the orthography of the language contributes to the differences in the prevalence (Landerl *et al.* 1997; Ziegler *et al.* 2010). For example, English is a deep orthographic language with no one-to-one correspondence of the letters and the phonemes, and could therefore result in more reading difficulties. On the other hand, a dyslexic child having a transparent orthographic language such as Finnish may manifest less severe deficits such as slow reading fluency (Holopainen *et al.* 2001; Puolakanaho *et al.* 2008; Serrano and Defior 2008).

2.2.4 Etiology

2.2.4.1 Heredity

The first documented observations of dyslexia having a tendency to cluster in families were made in the beginning of the 20th century by Thomas and Hinshelwood (Pennington 1990). These observations were later confirmed in 1950, when Hallgren conducted the first large-scale family study (Hallgren 1950). Since then numerous studies have supported the increased risk for dyslexia in relatives (Rutter and Yule 1975; Lewitter *et al.* 1980; Vogler *et al.* 1985; Pennington *et al.* 1991).

As family studies cannot assess the difference between genetic and environmental influences, twin studies can be used to separate these factors. In dyslexia, monozygotic twins have significantly greater concordance (68-100%) than dizygotic twins (38-54%), supporting the idea that there is a big genetic component in dyslexia (Bakwin 1973; DeFries *et al.* 1987; Harlaar *et al.* 2005; Hawke *et al.* 2006). The heritability estimate, which is a measure of the genetic component, spans from 0.37-0.72 for reading deficit (Willcutt *et al.* 2010). In support of this genetic influence is also the indications that children born to dyslexic families have an increased risk for language delay and later reading problems, which can be measured already during the first years of infancy (Viholainen *et al.* 2002; Lyytinen *et al.* 2004).

The inheritance pattern of dyslexia appears to be complex since there is no straightforward correspondence between a genotype and phenotype. Many factors such as genetic heterogeneity (distinct loci in different families), oligogeneity (allelic variants in multiple loci contributing to increased risk), incomplete penetrance (individuals with high-risk genotype but without dyslexia) and phenocopies (affected individuals without high-risk genotype) contribute to the complexity of the genetic nature of dyslexia (Fisher and DeFries 2002). However, in some families dyslexia appears to be transmitted as a single-gene defect; an autosomal dominant inheritance pattern has been seen in 20-30% of the families with dyslexic children (Pennington *et al.* 1991; Pennington 1995).

2.2.4.2 The male sex and environmental factors

Many studies show that dyslexia is more common among boys than girls (Berger *et al.* 1975; Katusic *et al.* 2001), although there is still debate if this is because of a bias in the diagnosis or if the male sex is a real biological predisposing factor. Evidence from a large scale epidemiological study of four large English-speaking populations suggests that boys are 2 to 3 times more likely to be affected with dyslexia than girls (Rutter *et al.* 2004). Moreover, gender ratios for reading difficulties are greater in more severely affected samples (Olson 2002). Sex-specific differences were also seen in one study analyzing variations in *DYX1C1* (Dahdouh *et al.* 2009).

Various social and environmental factors such as low birth weight, low parental education level and ethnic minority have been associated with an increased risk for dyslexia and school difficulties in general (Litt *et al.* 2005; Friend *et al.* 2008). Interestingly, the heritability of dyslexia is higher in families with high parental educational attainment than in families with lower educational level, suggesting gene and environmental interactions (G X E) (Friend *et al.* 2008). One interesting environmental factor that has been suggested to play a role in the development of reading is exposure to sex hormones during pregnancy. For instance, fetal testosterone levels during late pregnancy have been implicated in the development of neural structures important for reading (Geschwind and Galaburda 1985a; Geschwind and Galaburda 1985b; Beech and Beauvois 2006).

2.2.4.3 Comorbidities

Many neurodevelopmental disorders co-occur with each other more frequently than expected by chance indicating that there might be a common genetic basis for these different disorders. Dyslexic individuals often have problems in language acquisition, motor coordination and attention ability in addition to the reading and writing problems. It is therefore not a surprise that speech sound disorder (SSD), specific language impairment (SLI) and attention deficit hyperactivity disorder (ADHD) are frequent comorbidities with dyslexia (Gilger *et al.* 1992; Willcutt *et al.* 2000; Paracchini 2011). In the early studies, it was suggested that immune-related disorders and congenital cardiac anomalies occur in the families of dyslexics more often than by chance (Galaburda *et al.* 1985; Pennington *et al.* 1987).

2.2.5 Neurobiology

It is commonly acknowledged that dyslexia has a neurobiological basis. To understand this neurological origin of dyslexia, a range of investigations between dyslexic and normal readers have been conducted using post-mortem brain specimens, brain morphometric studies, functional brain imaging and electrophysiology methods.

2.2.5.1 Anatomical findings in dyslexic brains

Observations that dyslexic individuals would have defective development in the brain regions related to language such as left occipital or pariental lobes and angular gyrus were already made by Hinshelwood and Morgan (Richardson 1992). Later, Galaburda and colleagues made significant contributions to the field when they studied post-mortem brains of dyslexic individuals and found abnormalities in specific regions of

the brain. In the first study, they described abnormalities in the left cerebral hemisphere and posterior language area in a post-mortem brain of a 20-year-old man with severe reading disability (Galaburda and Kemper 1978). Moreover they have examined brains of four additional males (Galaburda *et al.* 1985) and three female dyslexic subjects (Humphreys *et al.* 1990). The main observations from these studies were that the brains of the dyslexic individuals had specific cortical malformations including ectopias (small pockets of misplaced neurons in the cortex) and microgyri (excessive number of convolutions in the surface of the brain), and absence of the usual left > right asymmetry of the planum temporale. The ectopias and malformations were mainly distributed across both frontal regions and in the left language areas. Myelinated scars were also found in the female patients who had lesser ectopias (Humphreys *et al.* 1990). The authors concluded that the anomalies seen in the post-mortem brains of dyslexics had a developmental origin, probably due to deficits in neuronal migration during the middle gestation. Although these findings are highly interesting, they are based on a very small number of individuals without unified diagnosis of dyslexia and should therefore be classified as initial discoveries.

To replicate the findings of Galaburda and colleagues, the magnetic resonance imaging (MRI) has been used to visualize and examine the brain morphology of dyslexic individuals. In contrast to the post-mortem studies, MRI can be used to study the brains *in vivo* and in larger sample sets with defined phenotypes. Most of the early MRI studies on dyslexic individuals confirmed the atypical symmetry (or reversed asymmetry) of the planum temporale (Hynd *et al.* 1990; Larsen *et al.* 1990; Kushch *et al.* 1993). However, some of the recent studies with more advanced MRI techniques did not confirm these findings (Leonard *et al.* 1993; Best and Demb 1999). Interestingly, when the brains of patients with periventricular nodular heterotopia (PNH) manifesting reading impairments were studied, similar disruptions of the cortical organization as in the post-mortem brain of the dyslexic individuals were found (Chang *et al.* 2005; Chang *et al.* 2007; Felker *et al.* 2011). This supports the hypothesis that disconnection of the cortical regions plays a critical role in reading fluency.

In addition, differences in the size and volume of corpus callosum, the largest white matter structure in the human brain connecting the two hemispheres, have been examined in dyslexic individuals. Increase size of the corpus callosum was found in individuals with dyslexia (Duara *et al.* 1991; Rumsey *et al.* 1996), but this was not supported in all the studies (Hynd *et al.* 1995; von Plessen *et al.* 2002). Nevertheless, some differences were found in all the studies leading to the conclusion that variation in the corpus callosum morphology could be associated with dyslexia. Furthermore, anatomical studies of dyslexic brains have suggested differences in the gray matter, cerebellum, white matter gyrification and total brain volume (Eckert *et al.* 2003; Casanova *et al.* 2004; Silani *et al.* 2005).

2.2.5.2 Brain activation patterns in dyslexic individuals during reading

The anatomical findings have demonstrated that there are differences between a "dyslexic brain" and a brain of a normal reader. If these suggested alterations in the brain structures (or others that have not yet been found) would have a functional consequence for reading and writing, this should be seen in the brain activation patterns

between the subjects under performance of reading related tasks. To test this hypothesis, functional neuroimaging techniques such as fMRI (functional magnetic resonance imaging) and MEG (magnetoencephalography) have been used to map the activation in different brain areas during reading related tasks.

Indeed, compiling evidence show that there is a specific "neural signature for dyslexia" that refers to the major finding from the functional imaging studies of significant differences in brain activation patterns between dyslexic and normal readers (Shaywitz and Shaywitz 2008). In general, the dyslexic readers have reduced activation in posterior system of reading processing mainly in the left parietal/temporal and left inferior temporal/occipital areas seen across different languages (Figure 2) (Salmelin *et al.* 1996; Paulesu *et al.* 2001; Temple *et al.* 2001; Shaywitz *et al.* 2002). In addition, some studies have shown a slight hyperactivation in the anterior region of reading (inferior frontal gyrus) during different reading related tasks (Shaywitz *et al.* 1998; Georgiewa *et al.* 2002). It has been speculated that this hyperactivation could serve as compensatory mechanism in the dyslexic children to cope with the reading tasks. This activation was also increased with age supporting this hypothesis (Shaywitz *et al.* 2002).

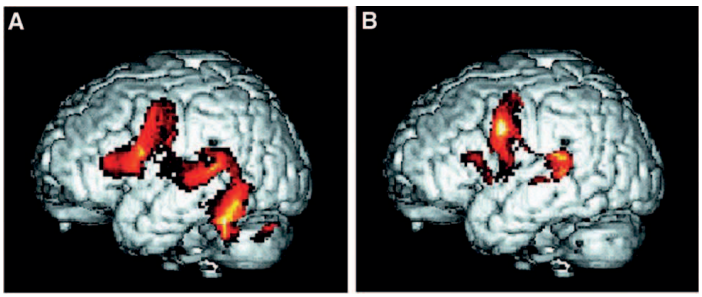

Figure 2 Differential brain activation patterns between normal (A) and dyslexic readers (B). *Science* (291) E. Paulesu *et al.* Dyslexia: Cultural Diversity and Biological Unity. Copyright 2001, with permission from AAAS.

In conclusion, both anatomical and functional differences have been found in the brain of dyslexic individuals compared to normal readers, but so far no single marker region or activation pattern in the brain can be used to diagnose children with dyslexia. This shows that dyslexia is a very complex phenotype also from the neurological point of view.

2.3 NEURONAL MIGRATION

The human brain is the most complex of all biological systems, as far as we know. The mature brain contains more than 100 billion neurons organized into different brain regions and structures (Herculano-Houzel 2009). The neocortex is the largest and most important information processing network of the brain that develops in the early fetal period containing approximately 20 billion neurons (Pakkenberg and Gundersen 1997). Most of the cortical neurons are generated during the early brain development and they migrate to their final positions in the neocortex where they then form connections that are essential for information processing. The migration of the newly born neurons is a highly regulated process and even slight alterations in this process can lead to problems in the information processing later in life. As it has been suggested that dyslexia could be caused by disturbances in neuronal migration, and since many of the candidate genes have been implicated to participate in this migration process, this chapter will summarize the normal neuronal migration process and what can happen when it is disrupted.

2.3.1 Normal neuronal migration

The development of the forebrain depends on two different modes of cell migration: radial and tangential (Figure 9A). Most of the neurons are produced in the ventricular zone (VZ) and migrate radially from the VZ to the developing neocortex. The other proliferative zone is located in the region of the ventral telencephalon, producing the cortical GABA(γ-aminobutyric acid)-ergic neurons that migrate to the cortex using tangential routes (Marin and Rubenstein 2001). Two major tangential routes have been identified: (1) from the medial ganglionic eminence (MGE) to the neocortex and hippocampus, and (2) from the lateral ganglionic eminence (LGE) to the olfactory bulb (Figure 3A) (Marin and Rubenstein 2001; Kriegstein and Noctor 2004).

The migration of neurons results in the formation of a six-layered structure of the mature cortex (Figure 9B) (Cooper 2008). This layered structure is formed in an "inside-out" manner, so that the first neurons migrating from the VZ to the cortex plate (CP) form the deepest layer VI. Then, the superficial layers are formed by newborn neurons migrating past the older neurons to reach their destinations in the CP. The neurons move to their position from the VZ in two different radial migration ways: somal translocation and locomotion (Figure 3B) (Kriegstein and Noctor 2004; Huang 2009). In the somal translocation, neurons extend a long leading process to the pial surface and then by shortening this process the cell body is pulled towards the pia. In contrast, in locomotion the entire cell moves guided by the radial glial cells. The migrating neurons usually have bipolar morphologies when they migrate. However, some neurons have complex multipolar morphologies in the lower parts of the developing neocortex. This multipolar morphology is a transient state of the neurons and they will eventually adopt a bipolar morphology to migrate. The transition to and from the multipolar stage is a sensitive period and deficits in this stage can cause disruption in the neuronal migration (LoTurco and Bai 2006). The positioning of the neurons into the correct layers is controlled mostly by the Cajal-Retzius cells (CR). These cells produce a molecular signal, called Reelin, that is part of the pathway that signals neurons when to stop migrating and take up their positions in cortex (Bielle *et al.* 2005; Huang 2009).

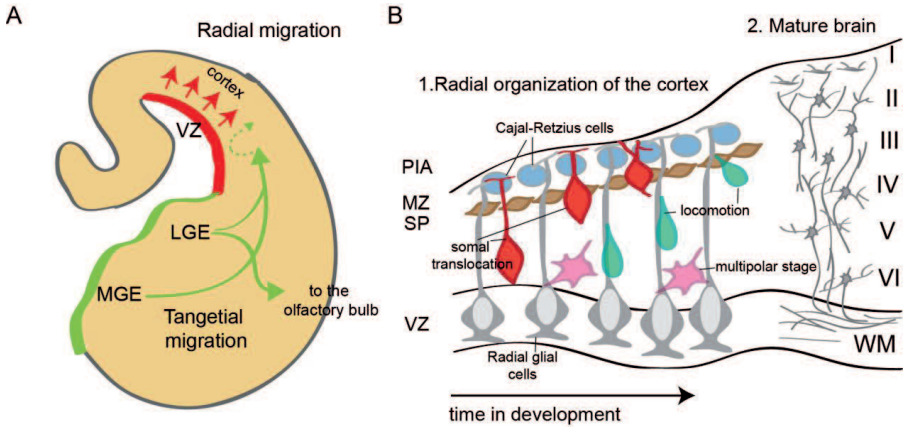

Figure 3 The neuronal migration routes and the organization of the neocortex (A) Routes of the radial (red) and tangential (green) neuronal migration from the ventricular zone (VZ) and medial- and lateral ganglionic eminence (MGE, LGE) to the cortex. (B) Different ways and organization of radial migration from the VZ to the subplate (SP) and marginal zone (MZ) containing Cajal-Retzius cells (1) to from finally the six-layer structure of the mature brain spanning from the white matter (WM) to the pia (2). Modified from *Cell* (128) Aylan *et al.*Trekking across the Brain: The Journey of Neuronal, Copyright 2007, with permission from Elsevier.

Once the neurons have reached their target region in the cortex, they need to develop neuronal processes, axons and dendrites, in order to integrate into neural networks. These processes allow them to communicate with other neurons.

2.3.2 Neuronal migration in disease

As summarized above, the normal development of the cortex requires proliferation and differentiation of the neural progenitor cells to neurons that then migrate to their functional layers in the cortex reaching from the white matter to the pial surface. Impairments in these steps can cause a variety of malformations that are called neuronal migration defects/disorders (NMD). The common deficits in the cortex formation and the corresponding malformations are summarized in the Table 1. NMDs also co-occur with many different syndromes such as Miller-Dieker Syndrome, Zellweger Syndrome and Walker-Warburg Syndrome (Liu 2011). Recent evidence suggests that subtle disturbances in the neuronal migration can cause many neurodevelopmental and psychiatric disorders such as epilepsy, schizophrenia and autism (Deutsch *et al.* 2010).

Table 1 Common cortical malformations and examples of the causal genes

DEFECT	MALFORMATION	GENE/S
Abnormal neuronal-glial proliferation or apoptosis	Microcephaly	*MCPH1, CDK5RAP2, ASPM, CENPJ*
	Megalencephaly	
Abnormal neuronal migration	Periventricular heterotopia	*FilaminA, ARFGEF2*
	Lissencephaly	*RELN, LIS1, DCX, ARX, YWHAE, VLDLR*
	Cobblestone cortex	*FCMD, FKRP, POMPT1, LARGE*
Abnormal cortical organization	Polymicrogyria	
	Cortical microdysgenesis	

Most of the NMDs have a genetic basis and causal mutations in many genes have been revealed (Table 1). The identification of genes causing severe malformations in the neocortex has also helped in understanding the molecular pathways needed for the normal neuronal migration such as microtubule-based transport, centrosome function and cell cycle pathway (Manzini and Walsh 2011; Pramparo *et al.* 2011). The most studied genes are *RELN,* LIS1 and *DCX* (*Doublecortin*) that all cause lissencephaly when mutated. Even before *RELN* was identified as a causative gene for lissencephaly, the *Reeler* mouse having a mutation in the *Reln* gene was a model for neuronal migration defects (D'Arcangelo *et al.* 1995). Extensive genetic, biochemical, and morphological studies have been carried out to determine the individual components of the Reelin signaling cascade. Reelin itself is a secreted extracellular matrix glycoprotein that can bind to the VLDLR and ApoER2 receptors to activate different signaling cascades such as phosphorylation of the intracellular adaptor molecule Disabled 1 (DAB1) (D'Arcangelo *et al.* 1999; Howell *et al.* 1999). LIS1 is connected to the Reelin pathway by interacting with the phosphorylated DAB1 (Assadi *et al.* 2003). However, the most studied interaction of LIS1 is with the motor protein dynein, both of which are required for the centrosome progression and nucleokinesis during neocortex development (Tsai *et al.* 2007). DCX has a well-established role in stabilizing microtubules and therefore it has been speculated that it controls the signaling-induced dynamics of microtubules in migrating and differentiating neurons (Gleeson *et al.* 1999; Schaar *et al.* 2004). In addition to the genes mentioned above, *DISC1* (*Disrupted in Schizophrenia 1*) is a new interesting gene involved in neuronal migration and many neuropsychiatric disorders (Brandon *et al.* 2009). Probably there are many more genes to be identified as causal genes for NMDs and more work to be done in order to fully understand all the pieces in the neuronal migration pathway.

2.4 THE HUMAN GENOME

Our genetic material, the human genome, consists of a large nuclear genome and a small mitochondrial genome. The nuclear genome comprises of 22 autosomal chromosome pairs and two sex chromosomes (X and Y). The genetic information is coded in the deoxyribonucleic acid (DNA) molecule composed of the nucleotide bases adenine (A), cytosine (C), guanine (G) and thymine (T). The DNA has a double helix structure that was discovered almost 60 years ago by James Watson and Francis Crick (Watson and Crick 1953). The double stranded DNA molecule is bound to histones and

coiled into nucleosomes and further to chromatin fibers to be packed and organized correctly in the nucleus.

A new era started within genome research when the human genome project (HGP) was launched in 1990 aiming to determine the sequence content and to identify all the genes in the genome. The first draft of the human genome was published in 2001 as a result of the HGP and the competing company Celera (Lander *et al.* 2001; Venter *et al.* 2001). It already comprised roughly 90 % of the total sequence. Now, ten years later a number the individual genomes have been published and efforts from the 1000 genomes project will provide even more detailed information of the genome, especially the variations in all aspects (Levy *et al.* 2007; Wheeler *et al.* 2008; 2010; Marth *et al.* 2011).

There are approximately 20-22,000 protein coding genes in the genome. However, the fraction of the sequence coding for those is only 1.1 %. Other conserved sequences such as RNA genes and regulatory sequences count for 4 % of the genome. This means that around 94.9 % of the nuclear genome consists of unique noncoding DNA and highly repeated sequences such as transposons and heterochromatin. The highly repetitive sequence was earlier considered as junk DNA, but recent evidence show that it can code for a number of RNA species and it may also contain important regulatory regions (Faulkner *et al.* 2009; Mattick 2011).

2.4.1 Sequence variations in the genome

When comparing the DNA sequence of several individuals, they are more than 99.7% alike. The manifestation of phenotypic differences such as height or risk for certain diseases can be partially explained by variations in the genome. The spectrum of variations spans from a single base pair difference to large chromosomal events. A single base difference can be called either single nucleotide polymorphism (SNP) or mutation. SNPs occur in every 100 to 300 bp and are present in at least 1 % of the population and mutations are rare events present in less than 1 % of the population (Figure 4A). SNPs constitute the majority of the genetic variation in the human genome (Consortium (2010).

Simple sequence repeats (SSR), micro- and minisatellites, are stretches of 1 to 6 nucleotide units repeated in a tandem form (Figure 4B). SSRs are very polymorphic due to the high mutation rate affecting the number of repeat units and comprise about 3 % of the genome (Lander *et al.* 2001). Both SNPs and SSRs have been very useful genetic markers when mapping loci for human disorders (*2.4*).

Structural variation (SV) is generally defined as a change in the chromosome structure greater than 1 kb in size. A large number of SVs consists of copy number variations (CNVs) where the number of genomic content is changed, such as insertions and deletions (Figure 4C). Inversions and reciprocal translocations are other forms of SVs (Figure 4C). The SVs can be either unbalanced or balanced, depending on whether there is a lost/gain of the genetic material or not. Now, when the sequencing of human genomes is becoming more routine, the spectrum of SVs and CNVs has widened to include much smaller events such as those <50 bp in length. (Mills *et al.* 2011)

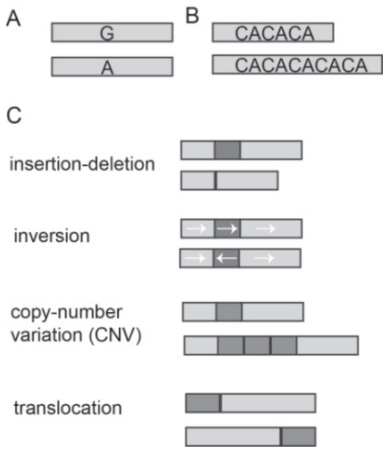

Figure 4 Sequence variations in the genome (A) Single nucleotide polymorphim (SNP) (B) Simple sequence repeat (SSR) (C) Different structural variations.

2.4.2 The central dogma of molecular biology

The central dogma of molecular biology refers to the flow of genetic information between DNA, RNA and protein. A specific DNA sequence in the genome, which is called a *gene,* consisting usually of multiple exons and introns, is used as a template to produce a functional complementary ribonucleic acid (RNA) sequence in the transcription process (Figure 5). The RNA molecules can be divided roughly in two classes: i) messenger RNA (mRNA) used as the information building block to make proteins and ii) noncoding RNAs (ncRNA) that are not translated into proteins. The ncRNA group contains RNA species such as the ribosomal RNA (rRNA), transfer RNA (tRNA), microRNA (miRNA) and long non-coding RNA (lncRNA), which have role in the ribosomes, translation process and carry out regulatory functions (Huntzinger and Izaurralde 2011; Wapinski and Chang 2011). The mRNA needs first to be spliced in a process, in which the introns are removed, then the sequence is read three bases at a time and the corresponding amino acid is added to build the protein (Figure 5). When the protein sequence is ready, post-translational modifications such phosphorylation and glycosylation further modify the sequence. Then finally the protein forms its structure and is ready to carry out its functions in cells (Figure 5).

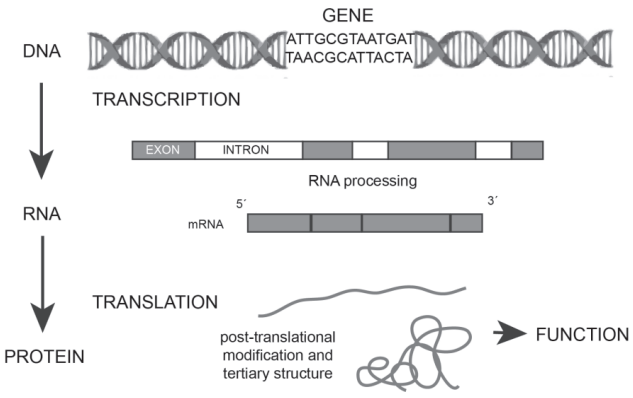

Figure 5 The flow of the genetic information from DNA to protein.

2.4.2.1 Regulation of gene expression

The human genome contains almost the same number of protein coding genes as the simple flatworm *Caenorhabditis elegans*. This means that the complexity of the organism obtained during evolution cannot be explained by gain or loss of genes. Instead, it has been suggested that the same set of genes can be used in different ways and that genes can achieve new biological functions (Carroll 2008). Thus, coordination of the spatial and temporal gene expression would serve as a mechanism to developmental and physiological complexity (Ravasi *et al.* 2010). For example, there are several hundreds of cell types in the human body that all contain the same genetic material; therefore to drive the differentiation and morphogenesis of each of the different cell types a tight monitoring of gene expression is needed. The regulation can take place in every step of the information flow from the transcription to post-modification and degradation of the proteins.

The regulation of gene expression is controlled by transcription factors (TFs) and their co-regulators binding to the transcription regulatory elements (TRE) located in the genome. The basic TRE of a gene is the promoter located usually upstream of the gene. For transcription to take place a transcription machinery consisting of the RNA polymerase II, TFs and co-factors are recruited to the promoter. The promoter can be the only TRE needed if the gene is needed all the time in the cells (a house-keeping gene). However for the other genes that have a complex and dynamic expression pattern (e.g. if they are needed in a specific time point during development or in a specific place), the regulation can be very complex. In that case, additional *cis* TREs that can act as enhancers or repressors when *trans*-acting factors are bound (MacQuarrie *et al.* 2011). These *cis* TREs can be located upstream, downstream and in the intronic regions of the gene, sometimes also very far away from the core promoter.

The identification of TRE for a specific gene is not trivial; usually the transcription factor binding sites (TFBSs) are short generic sequences and the *in silico* prediction can produce many false positives (Narlikar and Ovcharenko 2009). One way to identify

important regulatory elements in the genome is to compare the sequence conservation during evolution in the specific genomic region. Usually, the predicted TFBSs need to be experimentally verified with assays such as EMSA (*4.3.2*) and ChIP (*4.3.5.2*). The unraveling of regulatory elements in the genome and their use in different tissues will help in the task of finding disease genes and functional variants.

The regulation of gene expression can also happen in different other ways as pointed out. One of the hot topics is now epigenetics, where inherited DNA modifications, such as methylation and histone deacetylation, modify the gene expression without changing the DNA sequence itself. The epigenetics modifications may play a role in different diseases and also during development (Chahwan *et al.* 2011).

2.4.3 Beyond the central dogma

We already know how to read our genome and how the genetic information is transferred into function via RNA and proteins, but we are still far from understanding the complexity of the biological systems. The next step is to analyze complex biological information that we refer as different "-*omes*". -Ome stems from the Greek meaning complete or whole and "-*omics*" is a general term for studying a specific "-ome" such as genome, proteome, transcriptome and interactome (Figure 6). For instance, proteome refers to all the proteins in a certain sample e.g. a cell and different proteomic tools can used to study the proteome.

The main focus of the omics-field today is to map information from objects such as genes, proteins and ligands, finding interaction relationships between these objects and engineering networks to understand and manipulate regulatory mechanisms. Systems biology tools have been used in molecular medicine to for instance identify new diagnostic biomarkers and detect gene expression patterns. These tools can help to diagnose the severity of a certain disease, predict the disease outcome and identify the involved pathways. Particularly, this has been a successful approach in the field of breast cancer where distinct subtypes of breast cancer can be diagnosed using high-throughput microarrays (Perou and Borresen-Dale 2011). The systems biology approach has been useful for understanding of the complex nature of the biological systems and their interactions in health and disease (Figure 6). It is also a step closer to personalized medicine where the treatment is designed based on the genomic information from the patient combined with information of the disease pathways and pharmacokinetics (Ginsburg and Willard 2009).

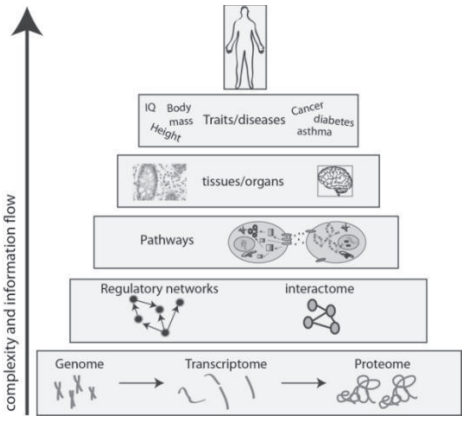

Figure 6 Complexity of the biological systems. The information to build biological systems is in the genome that is coding the transcriptome and proteome of a cell. The transcripts and proteins are connected in networks and pathways to carry out different functions in the cell. The different cell types form tissues and organs. Manifestations of traits and diseases come from the information in all these levels.

2.5 GENE DISCOVERY IN COMPLEX DISORDERS

The understanding of the genome content has expanded enormously during the last decade; still the identification of predisposing genes for a complex disorder is far from simple. This is due to many factors that affect the trait independently or together making the genetic studies and the interpretation of the results challenging. For example, 1) most of the complex disorders are inherited in a non-Mendelian manner, 2) there is no one to one correlation between the inherited genotype and phenotype (incomplete penetrance), 3) genetic heterogeneity and oligogeneity (distinct loci in different families or multiple genes contribute to the increased risk) and 4) environmental factors (Glazier *et al.* 2002).

The mapping of susceptibility genes is based on the use of polymorphic molecular markers in the genome (*2.3.1*). Today, SNPs are the most popular markers in mapping studies, since their availability increased dramatically after the HGP, the HapMap (www.hapmap.org) and the 1000 genomes (www.1000genomes.org) projects were released. Traditionally, the identification of the disease candidate genes have been done using positional cloning or by functional candidate gene studies. In positional cloning, the chromosome region containing the susceptibility gene/s for a certain disease is identified using genome-wide linkage analysis and further defined with higher resolution association analysis (Glazier *et al.* 2002). During the last five years the traditional positional cloning has been phased out by genome-wide association studies where the whole genome can be explored in a hypothesis-free manner with very high resolution (McCarthy *et al.* 2008).

2.5.1 Linkage and association analysis

Genetic linkage is based on the analysis of disorders in family pedigrees when a certain genetic marker is found co-segregating with the disorder phenotype. A parametric linkage analysis is based on a known disease model (when mode of heritance,

penetrance and frequency of disease allele in a population is known). However, when investigating a complex genetic disorder usually the disease model is unknown and non-parametric analysis is used. The hypothesis that the affected relatives would share a haplotype identical by descent (IBD) more often than the not affected in the region of a trait-causing gene, irrespective of the mode of heritance, is the basis of nonparametric linkage. The resolution of the linkage is not very good, however it can be used for genome-wide analysis to pin-point a certain chromosomal region that could harbor the susceptibility gene. (Borecki and Suarez 2001)

Association analysis aims to test whether allele frequencies differ between two groups, usually cases and controls. When looking at haplotypes (combination of alleles), the linkage disequilibrium (LD) structure of the genome is used to detect differences in cases and controls (Lander and Schork 1994). Association analysis has shown to be more sensitive to detect recessive disease alleles than linkage analysis, but it requires more markers to be examined (Risch and Merikangas 1996; Risch 2000). Additional difference between the methods is that linkage allows different alleles to be associated with the trait in different families, when in association analysis the allele that is associated with the trait is the same across population (Borecki and Suarez 2001).

2.5.2 Other approaches

2.5.2.1 Chromosomal rearrangements

Chromosomal abnormalities, such as balanced translocations, insertions and deletions that co-segregate with a disorder are valuable tools for identifying susceptibility genes (Bugge *et al.* 2000). The major advantage of using chromosomal rearrangements for mapping disease genes is that their position can be mapped exactly in the genome and therefore pin-point the affected gene. Of course, the rearrangements can also affect distant regulatory elements controlling genes even on different chromosomes. This can then make it difficult to distinguish the location of the true susceptibility gene. Mapping of balanced chromosomal translocations has been a successful approach in the search of dyslexia candidate genes (Taipale *et al.* 2003; Hannula-Jouppi *et al.* 2005). It must be noted that different kind of chromosomal rearrangements are present in most of the genomes without having any phenotypical consequences (Conrad *et al.* 2010).

2.5.2.2 Candidate gene studies

In the candidate gene approach, the gene to be tested is chosen by its position (positional candidate) or by its function (functional candidate). Positional candidate genes are usually within the region that was identified through linkage and association studies. If the pathology of the disease is known functional candidates can be studied based on their role on the disease pathway.

2.5.3 Genome-wide association studies

After the start of the HAPMAP project, the availability of common SNPs that can be used to capture the genetic variation in a population has increased enormously (2005; Frazer *et al.* 2007). This opened a new chapter in the field of genetics, especially for the complex diseases, when the genome-wide association (GWA) studies using up to million SNPs were started to map the disease loci in large sample sets. The threshold

for the GWA studies was set in 2007, when the Wellcome Trust Case-Control Consortium (WTCCC) published a paper in which they studied seven common complex diseases using 2000 cases and 3000 controls and identified 25 association signals with the $p < 5x10^{-7}$ (2007). In the beginning of August this year the GWAS database consisted of 1003 publications with association to 4978 SNPs (Hindorff). The GWAS approach has also been an excellent disease pathway finding tool, since many association signals to certain diseases come from genes in the same molecular pathways thus helping to understand the disease pathology and the designing of drug targets (Wang *et al.* 2010).

One of the bases for the GWA studies is the common disease/common variant (CD/CV) hypothesis. It assumes that the susceptibility factors for common diseases contain ancient common variants in the population (Reich and Lander 2001). Even though the GWA studies have been successful in identifying new susceptibility factors, there are limitations in this approach. One of the drawbacks has been that the identified susceptibility variants have only minor effects on the disease and therefore can account only a few percent of the trait heritability (McCarthy *et al.* 2008).

2.5.4 The future of gene mapping in complex disorders

Since the GWA studies have not succeeded in explaining all the genetic background of complex disease there is more work to be done. Especially within many neuropsychiatric disorders where very little of the genetic complexity has been revealed with the GWAS approach (Gershon *et al.* 2011). The concept of "missing heritability" or dark matter of genetics was launched after the results from the GWAS started to compile. This means that there are genetic mechanisms that were not taken into account in the CD/CV hypothesis. One popular explanation for the missing heritability has been that the GWA studies cannot capture rare and low frequency variants that have a MAF less than 0.05 (McCarthy *et al.* 2008). These variants could have larger effect sizes on the disease. Rare variants are known to cause Mendelian disorders but would maybe also explain part of the missing heritability of the complex traits, at least in some families. The low frequency variants (MAFs between 0.05 and 0.001) could have an intermediate effect on diseases and when taking into account gene-gene interactions the effect size could be even larger (McCarthy *et al.* 2008). Moreover, structural variants and inherited epigenetic modifications have already been shown to have effect on the genetics of complex disorders, but till remain understudied in many disorders.

At present, there is a new wave of genetic studies being published using next-generation sequencing of targeted regions, exomes and whole genomes to tackle the low and rare variants as well as structural variants in the genome. Since the cost of sequencing is decreasing all the time, this will be used to a larger extent during next years. The whole-exome sequencing has already proven to be a successful method in rare Mendelian traits, where the sequencing of only few unrelated individuals has been enough to identify the disease genes (Ku *et al.* 2011). However, many considerations need to be taken into account when studying complex diseases (as summarized in the beginning of this chapter); the sequencing approach is neither going be an easy task. Hopefully strategies such as the sequencing of carriers of extreme phenotypes in the

trait distribution or sequencing of multiple affected members of a family will result in interesting findings (Cirulli and Goldstein 2010).

Although association to a certain marker in a genome does not explain the pathogenesis of the disease, it can help with the diagnosis in the clinic. However, more needs to be done to understand the functionality of the associated markers or find the real causal markers, which are probably in LD with the associated one. Therefore, one of the major challenges in post-genomic research is to understand how physiological and pathological phenotypes arise from the networks or connectivity of the associated genes.

2.6 CANDIDATE LOCI AND GENES FOR DYSLEXIA

The strong genetic component of dyslexia has been shown by extensive family and twin studies. Use of molecular genetic methods for the mapping of dyslexia susceptibility loci in the human genome have provided convincing results, but have not been straightforward. So far nine chromosomal regions have been successfully mapped and assigned names from DYX1 to DYX9 (Table 2). These and some additional loci have been identified and confirmed in at least 20 independent linkage studies. Eight of these were genome-wide linkage screens specific for dyslexia samples sets and two for general reading and spelling ability. The remaining studies were replications of the highlighted regions from the genome-wide screens. Dyslexia candidate genes have been identified in some of the DYX loci with help of chromosomal aberrations and using high-resolution association studies, some of the genes and their functions will be discussed in the chapters below (Table 2).

It must be noted that the identification of susceptibility genes for dyslexia is not a closed chapter. Many more genes will surely be implicated in the etiology of this disorder when more samples sets are collected and new methods and approaches used.

Table 2 DYX# and additional loci identified from linkage studies with possible susceptibility genes identified in these regions. Genes in bold are the most replicated susceptibility genes.

LOCUS	MIM	Chr	Original publication	Replications	GENE/S	REFS
DYX1	127700	15q21	(Smith *et al.* 1983)	(Smith *et al.* 1991; Grigorenko *et al.* 1997; Schulte-Korne *et al.* 1998; Nothen *et al.* 1999; Morris *et al.* 2000; Chapman *et al.* 2004; Marino *et al.* 2004; Igo *et al.* 2006; Bates *et al.* 2007; Platko *et al.* 2008; Schumacher *et al.* 2008; Lewis *et al.* 2011)	**DYX1C1**, CYP19A1, ZNF280D, TCF12	(Taipale *et al.* 2003; Anthoni *et al.* 2011; Buonincontri *et al.* 2011)
DYX2	600202	6p22.2	(Smith *et al.* 1991; Cardon *et al.* 1994)	(Grigorenko *et al.* 1997; Fisher *et al.* 1999; Gayan *et al.* 1999; Grigorenko *et al.* 2000; Grigorenko *et al.* 2001; Fisher *et al.* 2002; Kaplan *et al.* 2002; Marlow *et al.* 2003; Turic *et al.* 2003)	**DCDC2, KIAA0319**, TTRAP	(Francks *et al.* 2004; Cope *et al.* 2005; Meng *et al.* 2005)
DYX3	604254	2p16-p11	(Fagerheim *et al.* 1999)	(Fisher *et al.* 1999; Fisher *et al.* 2002; Francks *et al.* 2002; Francks *et al.* 2002; Petryshen *et al.* 2002; Kaminen *et al.* 2003; Marlow *et al.* 2003; Peyrard-Janvid *et al.* 2004)	MRPL19, C2ORF3	(Anthoni *et al.* 2007)
DYX4	#127700	6q11.2-q12	(Petryshen *et al.* 2001)	(Bates *et al.* 2007)		
DYX5	606896	3p12-q13	(Nopola-Hemmi *et al.* 2001)	(Fisher *et al.* 2002; Marlow *et al.* 2003; Bates *et al.* 2007)	ROBO1	(Hannula-Jouppi *et al.* 2005)
DYX6	606616	18p11.2	(Fisher *et al.* 2002)	(Francks *et al.* 2002; Marlow *et al.* 2003; Bates *et al.* 2007)	PTPN2, C18OR15, MCR5, ZNF519	(Scerri *et al.* 2010)
DYX7	#127700	11p15.5	(Hsiung *et al.* 2004)	(Fisher *et al.* 2002; Marlow *et al.* 2003; Tzenova *et al.* 2004)	HRAS-DRD4	(Hsiung *et al.* 2004)
DYX8	608995	1p36-p34	(Rabin *et al.* 1993)	(Grigorenko *et al.* 2001; Tzenova *et al.* 2004)	KIAA0319L	(Couto *et al.* 2008)
DYX9	300509	Xq27.3	(de Kovel *et al.* 2004)	(Bates *et al.* 2007)		
		7q33	(Kaminen *et al.* 2003)	(Bates *et al.* 2007; Matsson *et al.* 2011)	DGKI,CHRM2, PTN, CRE2B3L2	(Matsson *et al.* 2011)
		18q12-q21	(Fisher *et al.* 2002)		C18ORF34, RIT2,DYM, NEDD4L	(Scerri *et al.* 2010)
		21q	(Fisher *et al.* 2002)		PCNT, DIP2A, S100B, PRMT2	(Poelmans *et al.* 2009)
		2q22	(Raskind *et al.* 2005)	(Bates *et al.* 2007)		
		13q	(Igo *et al.* 2006)			

2.6.1 DYX1C1

The main topic of this thesis and the first identified dyslexia candidate gene, *DYX1C1* (dyslexia susceptibility 1 candidate 1) was cloned with the help of a Finnish family in which a chromosomal translocation t(2;15)(q11;q21) co-segregated with dyslexia (Taipale *et al.* 2003). The family consisted of four translocation carriers: a father with history of reading and writing difficulties at school, two daughters both diagnosed with dyslexia and a son that had cognitive abilities below the normal range and severe problems in learning to read (Nopola-Hemmi *et al.* 2000). The dyslexia in the female carriers was based on deficits in verbal short term-memory and in phonological retrieval. The male proband of the family showed problems in verbal and visual memory, phonological awareness and retrieval, but he was not formally diagnosed with dyslexia since his IQ was just below the normal range.

The translocation disrupted the *DYX1C1* (then called EKN1) gene located near to the DYX1 locus (Table 2). *DYX1C1* was predicted to contain 10 exons giving the mRNA length of 1, 993 bp and encoding for a 420 aa protein (Taipale *et al.* 2003). The interval of the translocation included the exons 8 and 9 of the gene, and the breakpoint disrupted one of the tetratricopeptide repeat domains present in the protein (Figure 7). The putative promoter of *DYX1C1* was predicted to be located just in front of exon 1, but data in this thesis (III), ongoing work and databases (such as http://fantom.gsc.riken.jp/4/) suggests that the primary promoter used is situated in the exon 2 (Figure 7A).

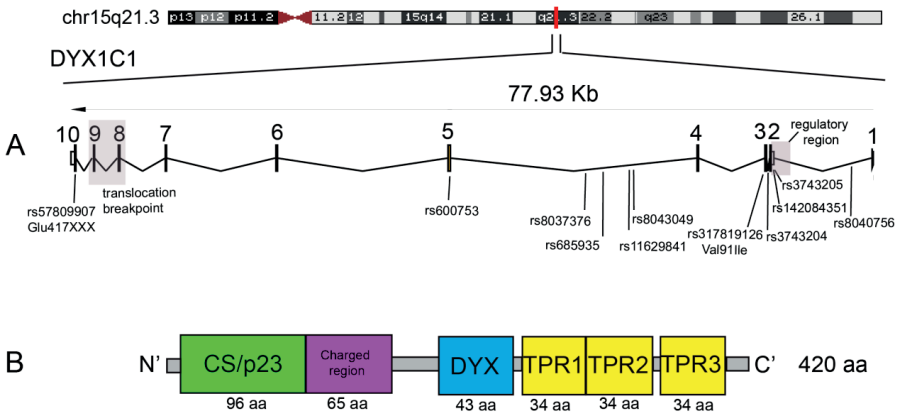

Figure 7 The chromosomal location and structure of DYX1C1. (A) Genomic structure and approximate location of the associated SNPs and the translocation breakpoint. (B) Protein structure of DYX1C1. Colored boxes represent different domains, TPR denotes tetratricopeptide repeat.

The protein domain structure of DYX1C1 is unique; the N-terminal part of the protein contains a p23 domain (also called the CS-domain) followed by a charged region (identified in Study IV), and the C-terminal part contains the DYX1-domain (identified in Study IV) and three tetratricopeptide repeat (TPR) domains (Figure 7B). The p23 domain in DYX1C1 is 96 residues long corresponding to the average size of this protein-protein interaction domain. The domain structure has a compact antiparallel beta-sandwich fold consisting of seven beta-strands (Garcia-Ranea *et al.* 2002). This

domain can be found in different eukaryotic protein families such as the SGT1 family, the nuclear movement nudC family and the p23/wos2 family. Interestingly, the NudC family contains the nudF protein with high homology to the human LIS1 protein. The conservation of the nuclear movement pathway of these proteins are thought to function in fetal brain development (Zheng *et al.* 2011). The TPR domain consists of a degenerate 34-amino acid repeated motif, which forms two antiparallel α helices packed into an open structure (Goebl and Yanagida 1991; Blatch and Lassle 1999). The TPRs can be found in many proteins that function in cell-cycle control, transcription repression, stress response, protein kinase inhibition, mitochondrial and peroxisomal protein transport and neurogenesis (Blatch and Lassle 1999). The function of the TPRs is to mediate protein–protein interactions and the assembly of multiprotein complexes. Interestingly, there is only one other protein that contain the p23 and TPRs in the human genome; SGT1, which is involved in kinetochore function and cell cycle (Steensgaard *et al.* 2004).

The identification of the DYX1-domain, located in front of the TPRs, is described in this thesis (5.4). The function of this domain is unknown and it has not been found in any other proteins in the databases. The DYX1C1 gene and protein sequences appear to be conserved during evolution ((Taipale *et al.* 2003) and Study IV) suggesting that it could have an essential role in basic cellular functions.

The expression pattern of *DYX1C1* has also been studied in different tissues demonstrating that it is abundantly present in brain, lung, kidney and testis (Taipale *et al.* 2003). The rat homolog *Dyx1c1* has been shown to be present during forebrain development with highest expression in the neocortex, hippocampus and choroid plexus (Threlkeld *et al.* 2007).

2.6.1.1 Association with dyslexia and different quantitative measurements

In addition to the identification of the translocation, Taipale and co-workers sequenced the coding parts of the DYX1C1 gene to identify polymorphisms in 20 unrelated dyslexic individuals (Taipale *et al.* 2003). A total of eight polymorphisms were found, from which two showed significant association with dyslexia in two sample sets of Finnish origin: the initial sample consisting of 35 dyslexics and 113 controls and the replication set of 54 dyslexics and 82 controls. The associated SNPs were also potential functional candidates: one introduces a stop codon truncating the protein by four amino acids (1249 G/T, rs57809907), and the other SNP was located in the 5'UTR (-3G/A, rs3743205), close to the translation initiation site affecting the gene expression (Figure 7A and Study I). Also a haplotype containing the -3A and 1249T alleles was significant when 9 trios were analyzed (Table 3).

Several groups have attempted to replicate the initial findings, but the results have been ambiguous. Table 3 summarizes all the association studies on dyslexia and related measurements using polymorphic variations in the *DYX1C1* to date. A total of 15 studies have tested the putative functional SNPs (rs3743205 and rs57809907). A significant association was found in seven studies, although the risk allele was not the same in all studies (Scerri *et al.* 2004; Wigg *et al.* 2004; Brkanac *et al.* 2007; Marino *et al.* 2007; Dahdouh *et al.* 2009; Lim *et al.* 2011; Newbury *et al.* 2011).

There are many explanations for the conflicting results of the association signals from *DYX1C1*. For instance the populations studied may have different haplotype structures and therefore show association to different alleles. Many of the case-control studies conducted to investigate the effect of these SNPs have been underpowered due to lack of samples to test. Interestingly, Dahdouh and colleagues demonstrated that variations in *DYX1C1* could have sex-specific genetic effects when they found an association only in the female probands with a haplotype spanning over the regulatory and coding regions of *DYX1C1* (Dahdouh *et al.* 2009). This suggests that sex should be taken into consideration when analyzing genetic studies of dyslexia. Recent results show that *DYX1C1* is also associated with general reading and writing ability, when variations in the gene were found to affect different reading and spelling measurements in two large population based sample sets from Australia (Bates *et al.* 2010; Paracchini *et al.* 2011). Interestingly, a significant association with short-term memory, which is a necessary component in reading, was also detected with one of the markers in *DYX1C1* (Nopola-Hemmi *et al.* 2000; Marino *et al.* 2007; Dahdouh *et al.* 2009; Bates *et al.* 2010). This is also in concordance with the short-term memory deficits seen in the translocation family (Nopola-Hemmi *et al.* 2000; Marino *et al.* 2007; Dahdouh *et al.* 2009; Bates *et al.* 2010). In addition, pleiotropic effects of the *DYX1C1* gene on mathematical traits were demonstrated in dyslexic families (Marino *et al.* 2011). Interestingly, most of the association signals are coming from the 5´region of the gene coding for the p23-domain (Figure 7A). However it may be that the "real" causal variations are still to be discovered.

Table 3 Association studies of DYX1C1 using dyslexia or population based sample sets. All the markers and haplotypes listed were concluded to be significant in the respective study. If nothing is noted the association was to dyslexia as a categorical trait.

Ref	Study population	Sample set	Measurements	No of SNPs studied	Single-marker association	Haplotype association
(Taipale *et al.* 2003)	Finnish	CC 109+195 Family 9	Dyslexia	8	rs3743205 rs57809907	rs3743205-rs57809907
(Wigg *et al.* 2004)	Canadian		Dyslexia Quantitative	6	rs11629841 rs3743205 *QM*	rs11629841-rs3743204 rs11629841-rs692691 rs11629841-rs692691 rs3743205-rs57809907
(Scerri *et al.* 2004)	UK	Family 264	Quantitative	8	rs57809907 *OC*	rs3743205-rs57809907 *OC*
(Cope *et al.* 2005)	UK	Trios 247	Dyslexia	3	NS	NS
(Marino *et al.* 2005)	Italian	Family 158	Quantitative	8	NS	
(Bellini *et al.* 2005)	Italian	CC 57+96	Dyslexia	3	rs142084351	
(Meng *et al.* 2005)	US	Family 150	Quantitative	2	NS	
(Marino *et al.* 2007)	Italian	Family 212	Quantitative	2	rs3743205 *SLBS*	rs3743205-rs57809907 *SLBS*
(Brkanac *et al.* 2007)	US	Trios 191 CC 191+192	Dyslexia	3	rs57809907	
(Saviour *et al.* 2008)	Indian	CC 52+51	Quantitative	4	NS	
(Dahdouh *et al.* 2009)	German	Trios 366	Dyslexia Quantitative	6	NS	Female proband: rs3743205-rs3743204-rs600753 *Dyslexia and STM*
(Bates *et al.* 2010)	Australian	Family 789 Population based	Quantitative	13	rs685935 *STM* rs171819126 *GM* rs3743204 *NW*	
(Newbury *et al.* 2011)	UK	Families 264 CC 331+336	Dyslexia Quantitative	2	rs57809907 *OC-choice*	
(Venkatesh *et al.* 2011)	Indian	CC 50+50	Dyslexia	6	NS	NS
(Paracchini *et al.* 2011)	Australian	Population sample set 520	Quantitative	25	rs8043049 *spelling* rs8040756 *reading* rs8037376 *spelling* rs7174102 *spelling*	
(Lim *et al.* 2011)	Chinese	Family 131	Dyslexia Quantitative	8	rs3743205 *Dyslexia and GM*	rs804075-rs3743205 rs3743205-rs4255730 rs692646-rs692691 *Dyslexia and GM* rs8040756-rs3743205-rs4255730

CC=case-control, GM= quantitative measurements, NW= non-word reading, STM= short- term memory, SLBS= single letter backward span, OC=ortographic coding.

2.6.1.2 *The role of Dyx1c1 in rat brains*

Functional studies using RNAi technique to knockdown the expression of rat *Dyx1c1* in a specific region of the brain during development have further strengthened the role of DYX1C1 in the etiology of dyslexia. Wang and colleagues showed that *in utero* knockdown of *Dyx1c1* arrested the migration of the transfected neurons close to their place of origin in the intermediate zone (IZ), subventricular zone (SVZ) and ventricular zone (VZ) (Figure 8A) (Wang *et al.* 2006). The control neurons migrated up to the cortex plate as expected (Figure 8B). The blocked migration due to the *Dyx1c1* knockdown could be rescued by overexpressing the full length *Dyx1c1* and a construct containing only the C-terminal TPR domains. Furthermore, expression of the *Dyx1c1* was necessary for the transition from the multipolar stage, which is a sensitive period in neuronal migration. Around 80 % of the Dyx1c1 knockdown neurons showed a multipolar morphology compared to 20 % in controls (Wang *et al.* 2006). This impairment of transition to and from the multipolar stage has also been seen in studies of other genes important for neuronal migration such as *Dcx, Flna* and *Lis1* (Bai *et al.* 2003; Nagano *et al.* 2004; Tsai and Gleeson 2005).

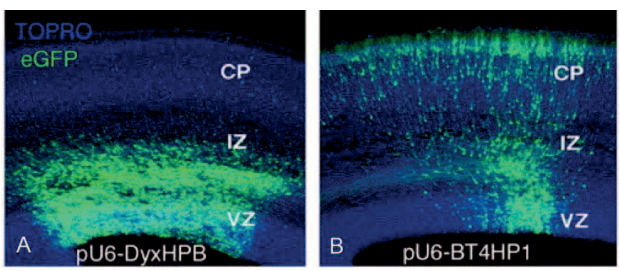

Figure 8 (A) The knockdown of Dyx1c1 in rats blocks the neuronal migration to the ventricular and intermediate zones (VZ and IZ). (B) In the controls the neurons migrate normally to the cortex plate (CP). Reprinted from Neuroscience (143) Wang et al. DYX1C1 functions in neuronal migration in developing neocortex. Copyright 2006, reprinted with permission from Elsevier.

Later, the impact of the impaired migration caused by the *Dyx1c1* shRNA was studied by examining the anatomical structure of adult rat brains (Rosen *et al.* 2007). This study revealed that the *Dyx1c1* shRNA resulted in neocortical and hippocampal malformations. The malformations found were very similar to those seen in the post-mortem brains of dyslexic individuals (Galaburda and Kemper 1978; Galaburda *et al.* 1985; Humphreys *et al.* 1990; Rosen *et al.* 2007). Disturbances in the hippocampus were seen in ~24 % of the *Dyx1c1* knockdown animals. These disturbances consisted of displaced cells from the pyramidal layer into the stratum radiatum and stratum oriens. The morphology of these misplaced neurons was more similar to neocortical pyramidal cells than to hippocampal cells. Interestingly, not all the displaced neurons were transfected with *Dyx1c1* shRNA and GFP suggesting secondary effects of the *Dyx1c1* shRNA in hippocampus. The ectopias were located in the neocortex and white matter. These clusters consisted of cells that were clearly neuronal but had disturbed radial orientation. Furthermore, the neurons distribution in the cortex differed between the *Dyx1c1* shRNA and control brains. The *Dyx1c1* shRNA brains had a bimodal distribution of the neurons; with 20 % of the neurons located close to the white matter and ~40 % very close to the pial surface instead of the Gaussian distribution seen in

controls. This indicates that even though a proportion of the shRNA *Dyx1c1* neurons do not migrate as normal, some of those neurons migrate beyond their target place. Additionally, there were a reduced number of neurons in the cortex structure in *Dyx1c1* shRNA rat brains compared to controls, probably due to an increase in cell death in this population of transfected neurons, since there was no change in the proliferation of the neurons (Rosen *et al.* 2007).

Recently, Currier and colleagues studied the suggested secondary effects seen in the brain malformations after the shRNA against *Dyx1c1* (Currier *et al.* 2011). They could confirm that the ectopias and heterotopias seen in *Dyx1c1* shRNA rats contained untransfected neurons. Also a large number of the neurons, such as GABAergic neurons, present in these clusters were born on E17.5 after the transfection and did not originate from the dorsal VZ place were injections were made. These results confirm that the embryonic knockdown of Dyx1c1 results in non-cell autonomous effects (Currier *et al.* 2011). Interestingly, the rescue with overexpression of *Dyx1c1* did not compensate these secondary effects in contrast to the rescue of the impaired migration (Wang *et al.* 2006; Currier *et al.* 2011).

It is difficult to assess a dyslexia phenotype in rats, but some behavioral tests can be used to study aspects of this complex phenotype such as learning and memory skills. When testing the adult rats having embryonic knockdown of *Dyx1c1* with different behavioral test, impairments in auditory processing and spatial learning were seen (Threlkeld *et al.* 2007; Szalkowski *et al.* 2011) An impairment in the auditory processing was seen in all of the rats in the *Dyx1c1* shRNA group, especially in the oddball task representing a more difficult auditory processing. This group did not show any improvement with experience. In addition, the rats that had hippocampal disruptions clearly showed a robust spatial learning impairment (Threlkeld *et al.* 2007). Further testing the rats using water maze tests, deficits in learning and memory were seen also persistent after 12 weeks of testing (Szalkowski *et al.* 2011). The different results on spatial learning and water maze tests suggested that there is a different circuit for hippocampal dependent spatial navigation ability. It must be noted that all the behavioral tests were conducted only in male rats.

2.6.1.3 DYX1C1 in cancer

There is also evidence that DYX1C1 could be important for development of different cancers. First, it was suggested that different splice variants of *DYX1C1* could be used as biomarkers for cancer more specifically colorectal cancer, since a differential expression was seen (Kim *et al.* 2009). Moreover, the *DYX1C1* was shown to be overexpressed in malignant breast cancer tissue (Chen *et al.* 2009). However, further studies are needed to address the role of DYX1C1 in cancer.

2.6.2 ROBO1

The *ROBO1* (roundabout, axon guidance receptor, homolog 1, also named as *DUTT1*) gene was identified with the help of a chromosomal translocation t(3;8)(p12;q11) in one individual with dyslexia. One of the translocation breakpoints was located in the DYX5 region and disrupted *ROBO1* between exons 1 and 2 (Nopola-Hemmi *et al.* 2001; Hannula-Jouppi *et al.* 2005). Sequencing of the coding regions of *ROBO1* in

dyslexic individuals from the large Finnish pedigree that showed linkage to the region (Nopola-Hemmi *et al.* 2001) revealed seven sequence variations but none of them were exclusively present in the affected. The specific haplotype associated with dyslexia in that family caused an attenuation of the mRNA expression of *ROBO1* suggesting that haplo-insufficiency of the gene may predispose to dyslexia (Hannula-Jouppi *et al.* 2005).

Until today no further association studies have been made to determine the role of *ROBO1* in dyslexia susceptibility, probably due to the lack of functional SNPs to test and the large genomic size of the gene. Recently, 144 tagging SNPs in ROBO1 were studied in 538 population based families for association with different makers of language impairment, memory and reading ability (Bates *et al.* 2011). They observed a significant association between multiple markers and language based measures such as phonological buffer capacity and short-term storage. Only one SNP reached the nominal significance for reading and spelling ability. These results suggest that ROBO1 is a strong candidate for the normal variation of language acquisition, supporting the earlier findings where the same locus has been linked to speech sound disorder (SSD) (Stein *et al.* 2004).

ROBO1 is an intriguing functional candidate for dyslexia for many reasons. It is a transmembrane receptor mediating the effects of SLIT proteins, which have important roles in the developing brain (Andrews *et al.* 2008). The Drosophila roundabout (robo) gene, which is an ortholog of ROBO1, has been shown to act as a gatekeeper of axonal crossing on the left-right axis of the brain in fruit fly (Kidd *et al.* 1998). Furthermore, it has been suggested that Robo1 is involved in cell migration, axon growth and guidance events (Andrews *et al.* 2006), and that it regulates the semaphoring signaling to guide the interneuron migration (Hernandez-Miranda *et al.* 2011).

2.6.3 DCDC2

The DYX2 locus on chromosome 6 is the most replicated susceptibility locus for dyslexia (Table 2) and a number of genes located in this region have been investigated and associated to dyslexia, among those is *DCDC2* (*Doublecortin domain containing 2*) (Deffenbacher *et al.* 2004; Francks *et al.* 2004). When the locus was studied with higher resolution, a significant association to *DCDC2* was found, but not for the other genes, supporting the role of *DCDC2* as the susceptibility factor in the locus (Meng *et al.* 2005; Schumacher *et al.* 2006). Since then, many studies have replicated the association to the *DCDC2* gene region with dyslexia and different quantitative measures (Brkanac *et al.* 2007; Wilcke *et al.* 2009; Marino *et al.* 2011; Scerri *et al.* 2011). Furthermore, association to the normal variation in reading and spelling was found with a marker in the *DCDC2* in a large population based family cohort from Australia (Lind *et al.* 2010). Interestingly, a region containing an associated short-tandem repeat marker (or deletion of the marker) in intron 2 appear to contain regulatory elements that can affect the gene expression (Meng *et al.* 2005; Meng *et al.* 2011). When the relationship between *DCDC2* and brain structure was analyzed significant changes were seen in the brain regions previously correlated with reading and symbol decoding (Meda *et al.* 2008). Furthermore, Meda and co-workers provided evidence that the deletion in the DCDC2 intron could affect the grey matter volume.

The *DCDC2* gene product belongs to a group of proteins containing tandem or single dcx domains. The *DCX* gene, the first characterized gene of the family, was identified after the discovery that mutations in the gene cause double cortex syndrome and lissencephaly (Gleeson *et al.* 1998). The knockdown of the rat homolog *Dcdc2* during development of the brain cortex results in similar phenotypes as for the knockdown of *Dyx1c1,* such as the blocked neuronal migration and anatomical disruptions in the postnatal cortex and hippocampus (Meng *et al.* 2005; Burbridge *et al.* 2008). Surprisingly, the *Dcdc2* knockout (KO) mouse does not display any disturbances in the brain structure implicating that it is not crucial for the normal neuronal migration (Wang *et al.* 2011). The authors speculated that the proteins of the DCX-family may have synergistic and redundant functions in brain development, since a more severe phenotype in the neuronal migration was seen in the *Dcdc2* KO mice after *in utero* knockdown of *Dcx* compared to controls (Wang *et al.* 2011). Additionally, cognitive abilities of the *Dcdc2* KO mouse were studied showing impairments in the visual discrimination and performance efficiency in a visuo-spatial learning and memory test (Gabel *et al.* 2011).

The cellular function of DCDC2 is mostly unknown, however the protein is known to localize along the microtubules. Recently, we linked the DCDC2 to primary cilia in neurons; more specifically overexpression of *DCDC2* increased the cilia length (Massinen *et al.* 2011). Furthermore, changes in *DCDC2* expression affected Sonic Hedgehog and Wnt signaling pathways known to be functional in cilia (Massinen *et al.* 2011).

2.6.4 KIAA0319

In addition to *DCDC2*, there is compelling evidence for another gene in the DYX2 locus, *KIAA0319*. The significant association with dyslexia and related quantitative measurements both in selected and population based sample sets have been detected, mainly to markers located in the 5´region of *KIAA0319* (Cope *et al.* 2005; Harold *et al.* 2006; Luciano *et al.* 2007; Paracchini *et al.* 2008). Indeed, a significant haplotype spanning the first four exons of the gene were shown to reduce the expression of *KIAA0319* (Paracchini *et al.* 2006). Furthermore a highly associated SNP in the haplotype creates a binding site for the transcription factor OCT-1 with functional consequences (Dennis *et al.* 2009). Interestingly, Couto and colleagues found evidence for an epigenetic based regulation of the *KIAA0319* expression in this haplotype region using ChIP against H3ac, a marker for accessible chromatin, to detect acetylated regions (Couto *et al.* 2010).

The protein coded by the *KIAA0319* gene contains a MANEC (motif at the N terminus with eight cysteines) domain, five polycystic kidney disease (PKD) domains and a Cys-rich domain (Velayos-Baeza *et al.* 2007). The protein is highly glycosylated and can be present in the plasma membrane, where it can be either secreted or internalized by the clathrin-mediated endocytosis pathway (Velayos-Baeza *et al.* 2008; Levecque *et al.* 2009; Velayos-Baeza *et al.* 2010). The role of KIAA0319 in neuronal migration has also been evaluated in the same manner as for DYX1C1 and DCDC2. This revealed that also the knockdown of *Kiaa0319* in rats reduces the distance migrated by the

neurons during cortex development (Paracchini *et al.* 2006). In contrast to the Dyx1c1 and Dcdc2 studies no "overmigration" phenotype was found, but instead the overexpression of the human *KIAA0319* also disrupted the normal migration of neurons (Peschansky *et al.* 2009). Hippocampal malformations were also found in the postnatal brain of rats subjected to shRNA against *Kiaa0319*, but to a lesser extent than in rats with shRNA against *Dyx1c1* and *Dcdc2* (Peschansky *et al.* 2009). In addition, the knockdown of Kiaa0319 resulted in morphological changes in the neurons, more specifically hypertrophy of apical dendrites.

2.6.5 Additional genes

2.6.5.1 Additional genes in the DYX1 locus

It has been suggested that the DYX1 locus may contain more susceptibility genes for dyslexia in addition to *DYX1C1*. First, there was another translocation identified in nearby the DYX1C1 translocation on chromosome 15 (Nopola-Hemmi *et al.* 2000) and second there is a controversy in the *DYX1C1* replication studies. Interestingly, the other suggested genes in the locus were also identified with help of translocations. Two genes *ZNF280D* and *TCF12* located 1.4 MB proximal to DYX1C1 were disrupted by a translocation t(6;15)(q22;q21) segregating with dyslexia in a Danish family (Buonincontri *et al.* 2011). Both are potential transcription factors, the *ZNF634* gene encodes for a C2H2-type zinc finger protein with homology to the Drosophila suppressor of hairy wing and the TCF12 is a member of the basic helix-loop-helix (bHLH) E-protein family. Furthermore, *CYP19A1* (cytochrome P450, family 19, subfamily A, polypeptide 1) has also been implicated to be a susceptibility gene in the locus, since the promoter region of the gene were disrupted by a translocation t(2;15)(p21;q21) in one individual with dyslexia and association to different language based disorders were found (Nopola-Hemmi *et al.* 2000; Anthoni *et al.* 2011). *CYP19A1* encodes the enzyme aromatase that converts androgens to estrogens. This makes it a highly interesting functional candidate gene since there is evidence that fetal hormones could affect the risk for dyslexia that is supported by the results presented in this thesis (II and III) (Geschwind and Galaburda 1985a; Geschwind and Galaburda 1985b; Beech and Beauvois 2006).

2.6.5.2 DYX3 locus: C2ORF3 and MRPL19

Two genes, *C2ORF3* (chromosome 2 open reading frame 3) and *MRPL19* (mitochondrial ribosomal protein 19) have been suggested to be susceptibility genes for dyslexia in the DYX3 locus after high resolution LD fine-mapping of the region using two dyslexia sample set of Finnish and German origin (Anthoni *et al.* 2007). MRPL19 belongs to the family of mitochondrial ribosome proteins, which are well characterized and important for oxidative phosphorylation. Interestingly, many of the mitochondrial ribosomal proteins have been associated with several neurological disorders (O'Brien *et al.* 2005). In contrast, *C2ORF3* is a novel gene of unknown function.

2.6.5.3 Multiple genes in the DYX6 locus

The linkage signal on the DYX6 locus spans over 40 MB along chromosome 18, and even though high-throughput SNP association analysis was performed the region could not be defined. In consistent with the linkage results, association signals were spread in

different genes in this region (Table 2) (Scerri *et al.* 2010). The authors concluded that the most consistent association signals were found in three genes *MCR5*, *DYM* and *NEDD4L*. The *NEDD4L* gene (neural precursor cell expressed, developmentally down-regulated 4-like) encodes a member of the NEDD4 family of HECT domain E3 ubiquitin ligases. It has a 78% amino acid sequence identity with *NEDD4*. NEDD4 is the most abundant E3 ubiquitin ligase in mammalian neurons and it is implicated in neurite growth and branching (Drinjakovic *et al.* 2010; Kawabe and Brose 2010). The *DYM* gene encoding for Dymeclin, a 669-amino acid protein of unknown function, causes the Dyggve-Melchior-Clausen dysplasia (DMC), a rare inherited dwarfism with severe mental retardation (Dimitrov *et al.* 2009). Dymeclin has been shown to shuttle rapidly between the cytosol and mature Golgi membranes and therefore it could be involved in cellular trafficking.

2.6.5.4 KIAA0319L in DYX8

Only gene studied so far in the DYX8 locus is the *KIAA0319L* gene. In addition to being a positional candidate gene it is a functional candidate by its homolog to *KIAA0319*. Couto and co-workers found a suggestive single marker association in the *KIAA0319L* with dyslexia and a haplotype association both with dyslexia and quantitative reading measurements (Couto *et al.* 2008). The locus contains also other interesting candidates, such as a homologous gene to *DCDC2* and SFPQ, which is involved the expression of DYX1C1.

2.6.5.5 Genes on chromosome 7

A region on the chromosome 7 was first identified to contain a susceptibility gene for dyslexia in a whole-genome linkage scan (Kaminen *et al.* 2003) and it was later replicated in an Australian samples set representing normal population (Bates *et al.* 2007). We further defined the 7q32-33 locus and identified significant overlapping haplotypes associated with dyslexia in Finnish and German dyslexia sample sets. The haplotypes are located in the region of the *DGKI (diacylglycerol kinase, iota)* gene which is a member of the type IV diacylglycerol kinase subfamily. The specific role of the enzyme encoded by this gene is undetermined; however it may play a crucial role in the production of phosphatidic acid in the retina or in recessive forms of retinal degeneration. In addition, SNPs in the gene have been associated to schizophrenia (Moskvina *et al.* 2009). Even though the significant overlapping haplotypes were in the *DGKI* gene, other genes in the region could also explain the association signal, such as *CHRM2* and *PTN* that are located close by. Cholinergic receptor, muscarinic 2 (*CHRM2*), is an acetylcholine binding receptor that has a role in working memory and synaptic plasticity (Seeger *et al.* 2004). Pleiotrophin (*PTN*) is a growth factor that has been shown to increase neurite outgrowth (Nakanishi *et al.* 2010) and play a role in learning and memory (Amet *et al.* 2001).

2.6.5.6 Genes on chromosome 21

The region on the chromosome 21 was first pointed out in a whole-genome scan using a dyslexia sample set from UK (Fisher *et al.* 2002). Even though there is no replication of the locus, four genes (*PCNT*, *DIP2A*, *S100B* and *PRMT2*) have been implicated as candidate genes in this region because they were disrupted by a deletion (Del21q22.3)

segregating with dyslexia in a family (Poelmans *et al.* 2009). All the genes are expressed in central nervous system (CNS) and are good functional candidates for dyslexia. For instance, pericentrin *(PCNT)* is expressed in the centrosomes and it is important for the normal functioning of the centrosomes and the cytoskeleton suggesting that it could have a role in migration (Matsuo *et al.* 2010). The DIP2A gene (also called *GRIP1B*) encodes the glutamate receptor interacting protein 1 variant b which may be involved in axon patterning in the central nervous system (CNS) (Yu *et al.* 2001). Interestingly, protein arginine methyltransferase 2 (PRMT2) has been shown to act as a co-activator for the estrogen receptor α and the androgen receptor (Qi *et al.* 2002; Meyer *et al.* 2007). The protein encoded by the *S100B* gene is a calcium-binding peptide and is used as a parameter of glial activation and/or glial death in many disorders of the CNS (Steiner *et al.* 2011).

2.7 ESTROGEN RECEPTORS AND THEIR SIGNALING

The work presented in this thesis connects the dyslexia candidate gene *DYX1C1* to estrogen signaling pathway (II and III). This chapter provides a brief introduction to estrogen receptors (ER) and their signaling, and some aspects of their distribution and function in the mammalian brain in order to understand the basis of the estrogen signaling and why it is interesting in the context of dyslexia.

ERs are members of the steroid receptor gene superfamily of nuclear receptors. There are two isoforms namely ERα and ERβ that show high degree of protein sequence similarity, mostly in the DNA-binding domain (DBD) and in the ligand binding domain (LBD). The ERs mediate the biological actions of estrogens, the most potent hormone being the gonadal steroid hormone 17β-estradiol (E2). E2 is involved in numerous physiological processes such as growth, differentiation, and function of the male and female reproductive tissues. In addition to E2, there are other ER ligands such as the commercial ligands propylpyrazol (PPT) a selective ligand of ERα and diarylpropionitril (DPN) ERβ agonist. (Dahlman-Wright *et al.* 2006)

The ERs can carry out their actions in at least two different ways in the cell using the genomic and the non-genomic pathways (Figure 9). In the genomic pathway, ERs dimerize upon ligand activation and translocate into the nucleus where it can bind directly to the estrogen response element (ERE) present in the DNA or via other TFs to regulate the transcription of target genes (Figure 9). It is well acknowledged that different co-regulators can affect the binding and regulation of the ERs (Katzenellenbogen *et al.* 2000). In the non-genomic pathway, the ERs can activate different protein-kinases cascades such as PI3K, MAPK and ERK cascades upon minutes after ligand binding (Figure 9) (Vasudevan and Pfaff 2008).

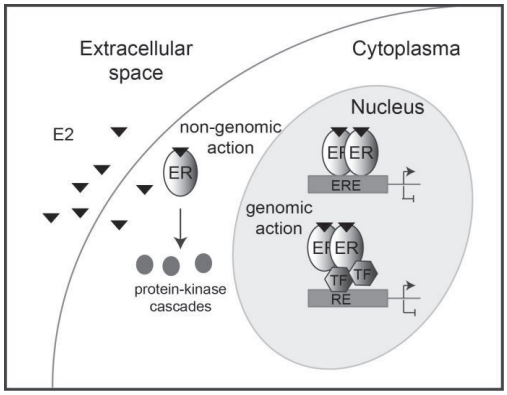

Figure 9 Estrogen receptor signaling. Upon ligand activation the estrogen receptor has two different action routes. In the genomic action, upon ligand activation ERs dimerize and translocate to the nucleus to activate or repress target genes by directly binding to the estrogen responsive element (ERE) or via other transcription factors (TF) and regulatory elements (RE). In the non-genomic action, ER activates different protein-kinase cascades in the cytoplasm.

2.7.1 Estrogen signaling in brain

Estrogen has multiple tasks such as to participate in neurogenesis, neuronal differentiation, and neuronal survival in the mammalian brain. To carry out these functions, both estrogen and its receptors should be present in specific brain areas. The highest levels of aromatase and ERs in brain are detected prenatally and during the first days after birth (McCarthy 2008). Estradiol in the fetal brain comes from the maternal circulation and placenta, but it can also be produced locally in the neurons by aromatization of the androgens to estrogens (Naftolin *et al.* 1988). Both ERs are present in the brain throughout development and into the adulthood, but the distribution of the two isoforms differs which suggest distinct functions.

During cortical development the expression of ERα precedes the expression of ERβ in the proliferative zones and it has shown to promote the proliferation of the progenitor cells (Martinez-Cerdeno *et al.* 2006). In humans, ERα can be detected at high levels throughout the cortex at gestational week 9, which is the time when the cortical plate appears (Gonzalez *et al.* 2007). ERβ expression in human embryonic brain cells was observed at gestational week 15, indicating that it plays a role later in development than ERα (Fried *et al.* 2004; Gonzalez *et al.* 2007). A direct evidence of the importance of ERβ in the brain development was demonstrated when studies showed that ERβ KO mice have neuronal migration deficits (Wang *et al.* 2001; Wang *et al.* 2003). The deficits were caused by abnormal migration of the neurons and an increased apoptotic neuronal death.

The protein levels of ERα are lower than ERβ in the human cortex and hippocampus during adult life (Osterlund and Hurd 2001). These brain structures are known to be needed for cognitive functions. Instead ERα is higher than ERβ in the areas of the rodent brain associated with reproduction, such as the hypothalamus (Osterlund and Hurd 2001). Both ERs are shown to affect learning: ERβ specifically influences hippocampus-dependent spatial learning, while ERα mainly affects emotional learning that is likely to rely on the amygdala (Fugger *et al.* 2000; Rissman *et al.* 2002). Furthermore, E2 has been suggested to have a protective role in schizophrenia and Alzheimer's disease. Estrogen has also very specific cellular roles in neurons such as promoting the neurite outgrowth and synaptic patterning (McCarthy 2008).

3 AIM OF THE THESIS

The overall aim of this thesis was to study the function and regulation of the dyslexia candidate gene *DYX1C1* to get insights on how it could be involved in the etiology of dyslexia. The specific aims were to:

I. Study the regulation of *DYX1C1* gene expression and whether this is affected by specific variations in the promoter and regulatory regions (I and III)

II. Determine if DYX1C1 is involved in estrogen signaling (II and III)

III. Identify the molecular networks of DYX1C1 (IV)

IV. Assess the role of DYX1C1 in cell migration using an *in vitro* assay (IV)

4 METHODOLOGY

This chapter briefly describes the main methods used in this thesis. All the detailed protocols for each experiment can be found in the respective papers.

4.1 CELL LINES

Cell lines derived from various tissues and species are useful tools in molecular biology research. They are important models for numerous diseases but also provide a valuable and ethical model for studying gene functions. Usually cell lines are easy to grow, proliferate indefinitely and exhibit a high degree of homogeneity making it easy to produce replicable results. However, there are some disadvantages when using cell lines; for instance, many of them are derived from tumors and therefore their genome can contain many chromosomal re-arrangements and mutations. Continuous culturing of the cells can produce phenotypic differences, thus subpopulations of cells within the cell line may arise making it harder to compare results from different laboratories. This problem has been discussed in the case of the MCF-7 cell line (Osborne *et al.* 1987). Below is a short summary of the cell lines used in this thesis and their origin.

The SH-SY5Y cell line is thrice-cloned from the SK-N-SH cell line which in turn is derived from a human neuroblastoma (Biedler *et al.* 1978). The cell line is genetically female and has trisomy of 1q. SH-SY5Y cells are described as neuroblast-like and they can be differentiated to neuron-like cells (Pahlman *et al.* 1984). We have used the SH-SY5Y cells for most of our experiments since it is a well-studied and often used cell line for studying neuronal biochemistry. The SH-SY5Y^{DYX1C1} cell line that overexpresses the full-length *DYX1C1* was constructed by transfecting the plasmid containing the *DYX1C1* gene into the SH-SY5Y cell line followed by antibiotic selection with Geneticin. Colonies were chosen and expanded in a selective media and DYX1C1 expression was confirmed by qRT-PCR and western blot (IV).

The human luminal epithelial ductal breast cancer cell line MCF-7 has a high endogenous expression of ERα which makes them responsive to estrogen. The MCF-7 is the most common cell line utilized in the estrogen receptor signaling studies; therefore we used it to study the involvement of DYX1C1 in ER signaling in Studies II and III. We also used the COS-7 cell line, kidney fibroblast cells derived from African green monkey in Study II.

In Study II, primary rat hippocampal neuron cultures were prepared as previously described (Brewer and Cotman 1989). The advantage of using primary cells is that they are taken directly from a living tissue, in this case from brains of E17 rat fetuses. These cells have undergone few population doublings in culture; therefore they are more representative of functional elements of the tissue that they were derived from than tumor or immortalized cell lines.

4.2 GENE EXPRESSION ANALYSIS

4.2.1 Quantitative real-time polymerase chain reaction (I, III, IV)

Quantitative real-time polymerase chain reaction (qRT-PCR) is a method for measuring gene expression. It is based on the detection of fluorescence produced by reporter molecules; two of the most common methods are SYBR Green I dye and Taqman® double-dye probes. SYBR Green I dye is a non-specific detection method where the dye binds to double-stranded DNA. On the other hand, Taqman® is a specific detection method using hybridization probes containing a fluorophore and a quencher group, designed to match the gene of interest. In qRT-PCR, the amount of amplification product is measured in "real-time", every PCR cycle yields a quantitative measurement of the PCR product. For relative quantification of the PCR products in qRT-PCR, the most common method is the $2^{-\Delta\Delta C}{}_T$ in which the C_T –value is the number of cycles needed to reach the arbitrary threshold (Livak and Schmittgen 2001). For the calculations, C_T -values of the gene of interest and a housekeeping gene are needed; the subtraction between those two is then ΔC_T -value, which is then used to calculate $\Delta\Delta C_T$ by using one sample as control and comparing all the other samples to that. The negative value of $\Delta\Delta C_T$ is then used as the exponent of two and represents the relative differences between the samples. The exponent conversion used in the $2^{-\Delta\Delta C}{}_T$ method is based on the assumption that the PCR is 100 % efficient e.g. the PCR product amount is doubled in every cycle.

Both SYBR and Taqman® based assays were used in the work described in this thesis. For the SYBR assays, melting curve analysis have been conducted for each primer pair to confirm a single PCR product and avoid non-specific products to generate fluorescent signal.

4.2.2 Expression microarrays (IV)

High-density oligonucleotide microarrays can be used to measure the expression of multiple genes simultaneously such as in whole-genome expression analysis of a complex RNA sample. The microarray slides contain DNA oligonucleotide probes designed to correspond to the studied genes. The samples containing fluorescent tags are hybridized to the slide and the intensity of the fluorescence is measured giving the read out of the gene expression in the sample. Since microarrays contain many thousands of probes, there is the possibility of cross-reactivity of samples to different genes giving then a false estimation of gene expression annotated for that probe. Illumina BeadArray™ technology uses probes that are more than 70 nucleotides long consisting of a 25-nucleotide code for identification and a 50-nucleotide gene-specific sequence. The arrangement of probes in the Illumina arrays is random and the positions are determined later.

In Study IV, we have used the Illumina's HumanHT-12 v4 Expression BeadChips, which contain more than 47,000 probes, to measure the whole-genome expression of SH-SY5Y, SH-SY5Y^{DYX1C1} and SH-SY5Y cells transfected with siRNA (siDYX1C1 and siControl). Hybridizations and scannings were carried out according to manufacturer's protocols (www.illumina.com). Pre-processing, normalization and statistical analysis of the microarray data was done using R statistical program with

implementation of the relevant packages lumi and limma (Smyth 2004; Du *et al.* 2008). To determine fold changes and standard errors, each probe was fitted to a linear model and empirical Bayes smoothing was applied to standard errors. Significantly perturbed probes for the two contracts, SH-SY5Y - SH-SY5Y^{DYX1C1} and siDYX1C1- siControl, were selected based on the adjusted p-values ($p < 0.01$) and B statistics ($B > 1$). B-value is a log odds probability score for the differential expression of the probe.

4.3 PROTEIN ANALYSIS

4.3.1 Western blot (I, II, III and IV)

The western blot method is used to study protein expression levels with antibody detection. It can also be used to verify the specificity of a given antibody. First, denatured proteins are separated by their molecular weight in sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and then transferred to a membrane. The protein of interest is then subsequently detected using a specific primary antibody and a secondary antibody conjugated with an enzyme for detection such as horseradish peroxidase (HRP). In this thesis, western blots have been used to measure differences in protein expression levels of DYX1C1, ERα, ERβ, TFII-I and c-MYB (II and III). We have also used the method to test the presence of proteins after Co-IPs (see 4.3.5.1) and show specificity of the used antibodies (II and IV).

4.3.2 Electrophoretic mobility shift assay (I)

An electrophoretic mobility shift assay (EMSA) is used to study protein-DNA interactions. In the assay, a duplex radiolabeled oligonucleotide probe and protein sample are mixed and separated in a polyacrylamide gel. In Study I, we analyzed protein binding affinities to different alleles in the promoter and regulatory regions of the *DYX1C1* gene. Double stranded DNA probes consisting of 30 nucleotides containing each of the alleles of rs3743205, rs12899331 and rs16787 were synthesized and radiolabeled. The probes were incubated with nuclear or whole cell extracts of SH-SY5Y cell line. In addition, we performed competition assays with a commercial consensus probe for the SP1 binding site and a serum response element (SRE) probe that is known to bind to TFII-I in the c-fos promoter confirming the identity of the bounded proteins.

4.3.3 Immunocytochemistry (I, II, III and IV)

Immunocytochemistry (ICC) is used for detection and visualization of proteins in cultured cells. The specific antibodies target peptides or protein antigens in the cells followed by detection with different methods. We have used an indirect method, where the primary antibodies are detected with secondary antibodies conjugated either to substrate (such as biotin) or to fluorophore (such as FITC). In Studies I and II, we used ICC with fluorophore detection to determine the subcellular localization of DYX1C1 and to study the co-localization of the exogenous DYX1C1 with exogenous and endogenous ERs proteins in cell lines. In paper III, we detected the endogenous DYX1C1 and ERβ in the SH-SY5Y cell line with biotin-avidin detection.

4.3.4 In situ proximity ligation (II, IV)

In situ proximity ligation (isPLA or in situ PLA) is used to study and visualize endogenous protein-protein interactions. In the isPLA method, first two primary antibodies raised in different species detect the protein/proteins of interest, and then secondary antibodies conjugated with unique oligonucleotide probes bind to the primary antibodies (Figure 10A and B). When the probes are proximal to each other they can be linked with addition of two oligonucleotide probes and an enzymatic ligation. The joined probes are then be used as a template for rolling circle amplification (RCA) of DNA using a polymerase (Figure 10C). After the amplification, the produced replication of DNA circle are visualized by labeled complementary oligonucleotide probes giving a fluorescent signal (Figure 10D). (Fredriksson *et al.* 2002; Soderberg *et al.* 2006)

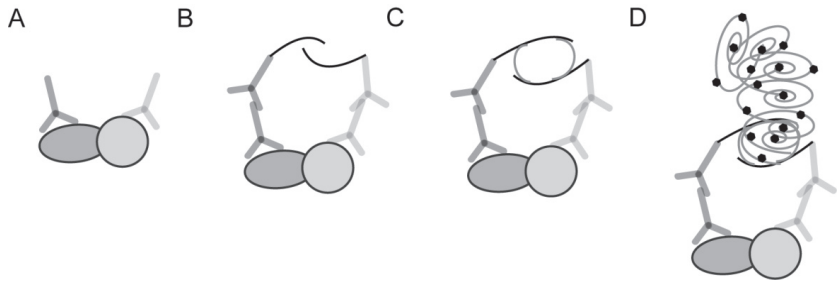

Figure 10 Schematic presentation of the in situ proximity ligation method (isPLA). (A) Two proteins in close proximity to each other are detected with corresponding antibodies raised in different animals. (B) Secondary antibodies with conjugated oligonucleotide probes are then attached to the primary antibodies. (C) The probes are then linked with addition of two more probes that then form a circle. (D) The circle produced by the probes are used as a template for rolling circle amplification (RCA) using a polymerase which is then detected by fluorescent labeled probes.

In Studies II and IV, we have used isPLA to verify the protein-protein interactions between DYX1C1 and ERs, DCDC2 or LIS1. We also studied the localization of DYX1C1-ER interactions in primary rat hippocampal neurons (II).

4.3.5 Immunoprecipitation (II, III and IV)

4.3.5.1 Protein complex immunoprecipitation (II and IV)

In the Co-immunoprecipitation (Co-IP) method, an antibody is used to immunoprecipitate the protein of interest together with a possible protein interacting complex from a cell lysate. Following the Co-IP, the protein-protein interactions partners can be detected by a western blot (4.3.1) using an antibody against a candidate interacting protein. It is also possible to detect novel interacting proteins by using a more high-throughput method such as LC-MS/MS (4.3.6). It is important to validate the Co-IPs with different proteins in the complex. We have used targeted Co-IPs to study the protein-protein interactions between DYX1C1, ERs and CHIP in Study II, and in Study IV to study the interactions between DYX1C1 with DCDC2, KIAA0319 and LIS1. We also used a global scale identification of the DYX1C1 interacting partners using Co-IP followed by LC-MS/MS (IV).

4.3.5.2 Chromatin immunoprecipitation (III)

Chromatin immunoprecipitation (ChIP) can be used to study the DNA binding sites of proteins, such as TFs or histone proteins. A specific antibody against the DNA binding protein is used to immunoprecipitate the protein of interest and together with associated DNA sequence. Before the immunoprecipitation the cells/tissues are cross-linked usually with formaldehyde and then lysed followed by sonication of the DNA. After the immunoprecipitation, the identification of the DNA regions present in the sample can be tested by PCR, qRT-PCR or by high throughput sequencing. A re-ChIP experiment can be performed to investigate if two proteins are present at the same site of the genome at the same time. In re-ChIP, two different antibodies are used sequentially in the immunoprecipitation step.

In Study III, we examined the binding of ERα, ERβ, TFII-I, c-MYB and RFX1 to a genomic region upstream of the DYX1C1 translation start site. Re-ChIP assays were also performed for the ERβ - TFII-I, ERβ – c-MYB and TFII-I – c-MYB pairs. The detection of the region studied was done by conventional PCR and qRT-PCR.

4.3.6 Protein identification by LC-MS/MS (I, IV)

Mass spectrometry (MS) is a commonly used analytical method where mass/charge ratio (*m/z*) of ions is measured in a vacuum. The molecular masses can be determined from the MS data, which then allows the molecular composition of a given sample to be resolved. In this thesis, we have used MS to identify proteins in different samples. Usually when analyzing proteins with MS, the analyte is a collection of peptides derived from the protein sample by digestion with e.g. trypsin. A separation of the digested peptides can be done with liquid chromatography (LC) prior to the MS analysis. After the MS analysis, the final identification of the protein is done by correlative database searches of the obtained peptide masses or to derive *de novo* sequences that can be used for standard queries in similarity search algorithms based on BLAST and FASTA.

Liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS) was used in Studies I and IV. In Study I, we identified proteins associated to the region containing the rs3743205. In Study IV, we identified novel interacting proteins of DYX1C1 by Co-IPs (*4.3.5.1*) combined with LC-MS/MS.

4.4 LIVE-CELL IMAGING

Time-lapse video microscopy of SH-SY5Y cells was used to analyze the random cell migration after perturbations of DYX1C1 levels (IV). The changes in *DYX1C1* levels were performed by transient transfection of siRNA and expression plasmids containing DYX1C1 or domain deletion constructs coupled with GFP. Twenty-four hours after the transfections, the cell media was changed and the cells were transferred for imaging to a Revolution system from ANDOR Technology, plc., UK using an Olympus IX81 microscope, a Yokagawa spinning disk and an Andor Ixon987 EMCCD camera. Bright field and fluorescence images were acquired every 10 min for 6 h using air objectives (10x NA 0.4). To obtain an effective field containing as many cells as possible, we acquired 25 fields in each well which were then stitched together as a single image.

Cell tracking was performed using the Time-lapse analyzer (TLA) program (Huth *et al.* 2010). The statistical significance of the results was analyzed by student t-test or ANOVA followed by tukey's post hoc test.

4.5 LUCIFERASE ASSAYS

Luciferase reporter assay can be used to test the effects of different DNA sequences on gene expression. The studied DNA sequence is cloned into an expression plasmid containing the luciferase gene that can be transfected into cells. The produced luciferase can then be measured then as light using a specific chemical reaction and a luminometer. In Studies I and III, we used luciferase assays to test if different SNPs located in the regulatory regions of *DYX1C1* could affect the gene expression. We also used a luciferase assay to study the ER signaling. For this we used an expression plasmid containing the estrogen receptor binding site (ERE) in front of the luciferase gene (II).

4.6 BIOINFORMATIC TOOLS

A selection of bioinformatic tools was used in the studies to generate and test hypotheses and to extract information from gene lists. We used the transcription element search system (TESS; http://www.cbil.upenn.edu/cgi-bin/tess/tess) and the Genomatix package (http://www.genomatix.de) to analyze the properties of nucleotide variants in the regulatory regions of *DYX1C1* (I) and to identify TFBSs for ChIP analysis (III).

Multiple sequence alignment of DYX1C1 was done to study the conservation of the protein across species and identify functional elements in the protein (IV). This was carried out using Clustal_X and Muscle (Thompson *et al.* 1997; Edgar 2004). The sequence information was also used to reconstruct a Neighbor joining tree showing the relationship between sequences of DYX1C1 protein from different species, this was carried out using Clustal_X with 100 bootstrap trials.

In Study IV, we produced a number of lists containing tens to hundreds of gene IDs. To extract the relevant biological information from the gene lists, we analyzed the gene ontology (GO) and KEGG biological pathway enrichment. For this purpose, we used the WebGestalt (http://bioinfo.vanderbilt.edu/webgestalt/) (Zhang *et al.* 2005; Duncan *et al.* 2010) and DAVID (http://david.abcc.ncifcrf.gov/) (Huang da *et al.* 2009; Huang da *et al.* 2009) bioinformatics resources. These web tools were used with default parameters and human genome or the whole gene set in Illumina Human HT-12 arrays as a background for the analysis. Adjusted p-value 0.05 and minimum of three genes were used as a threshold for significance.

5 RESULTS AND DISCUSSION

5.1 THE REGULATION OF *DYX1C1* EXPRESSION (I AND III)

As reviewed in the background section, a number of genes have been genetically linked to dyslexia (Table 2). However no direct causal variants or mutations have been found in these genes that could give a simple explanation for the association. It has been suggested that altered expression of the disease susceptibility genes, and not coding changes disrupting the gene, could be the cause of the increased risk (Cookson *et al.* 2009). Therefore, we have studied the regulation of *DYX1C1* expression, especially the effects of the associated polymorphisms in the regulatory region of the gene.

In Study I, we characterized three SNPs (rs1289331, rs16787 and rs3743205) located in the promoter and 5´UTR of *DYX1C1*. These SNPs were chosen since they form a haplotype associated with dyslexia (Dahdouh *et al.* 2009). It was predicted that the different alleles of these polymorphisms could alter the affinity of TFs binding to the corresponding DNA sequences by *in silico* tools. For instance, TFII-I and ELK-1 were predicted to bind to the A allele of the rs3743205 (also called the -3A), but not to the G allele. We used EMSAs and luciferase reporter assays to detect possible differences between the alleles of all three SNPs. Indeed, we observed allele-specific differential retardation for each of the three SNPs studied in EMSA. For the rs16787 and rs3743205 alleles, the differential retardation was very clear: rs16787C and -3G alleles caused an extra binding of protein/s. We could also show that the difference was due to the predicted factors SP1 family and TFII-I by competition assays. In the case of rs1289331 the differential binding was lesser although an intensity difference was seen between the alleles; a stronger binding of factors was seen to the T allele. We could also reduce this binding by a competition assay blocking the SP1 transcription factor family. In addition, we used luciferase reporter assays for the functional verification of the allele-specific binding differences. Again a clear significant difference between the alleles of rs16787 and rs3743205 was demonstrated and a smaller less significant difference between the rs1289331 alleles, confirming the EMSA results. Our results indicate that the studied polymorphisms in the regulatory region of DYX1C1 are functional and provide a mechanistic effect that may explain the association signal.

Furthermore, we observed that TFII-I bind together with SFPQ and PARP1 to the genomic sequence around the rs3743205 using protein identification by LC-MS/MS. We also concluded that the TFII-I, PARP1 and SFPQ increased the expression of *DYX1C1* measured by qRT-PCR. The regulatory effect of this protein complex was more efficient when -3G allele was present in the *DYX1C1* sequence in contrast to the -3A allele. TFII-I, PARP1 and SFPQ are highly interesting genes in the context of regulating a dyslexia candidate gene, and their function is discussed in chapter *5.1.1*.

In addition, we analyzed the regulation of *DYX1C1* expression in Study III. We explored if addition of 17β-estradiol (E2), a steroid hormone, could modulate the levels of *DYX1C1* in SH-SY5Y and MCF-7 cell lines. A rapid effect on the endogenous *DYX1C1* mRNA levels was seen after 4 h treatment of E2 in SH-SY5Y cells. The

protein levels of DYX1C1 were also increased after addition of E2 and DPN, a specific ERβ ligand.

To explain this E2-dependent regulation mechanistically, we did a detailed ChIP analysis of the possible binding of ERα and ERβ to the regulatory regions of DYX1C1. No binding was detected to the ERE half sites found in the promoter region, instead an E2 dependent binding of ERβ was found to the 5' UTR of *DYX1C1*. As there were no known ERβ binding sites in this regulatory region of *DYX1C1*, we investigated which other factors are recruited and which of those could mediate ERβ binding. In addition to TFII-I, two other TFs were predicted to bind; c-MYB (myeloblastosis viral oncogene homolog) and RFX1 (regulatory factor X1). We confirmed the binding of all these factors by ChIP. Interestingly, a clear enrichment of ERβ and TFII-I binding was detected upon E2 treatment. c-MYB was only recruited in the control treated cells while RFX1 was recruited both in E2 and control treated cells. Thus, ERβ and TFII-I, but not c-MYB or RFX1, appear to bind the regulatory region of *DYX1C1* in an E2 dependent manner. Re-ChIP analysis revealed that TFII-I and ERβ are present simultaneously after E2 treatment, whereas c-MYB is not recruited together with ERβ or TFII-I.

Furthermore, we analyzed the E2 dependent regulatory effect on *DYX1C1* using siRNA against ERβ, TFII-I and c-MYB. By this means, we could determine that both ERβ and TFII-I are needed for this regulation, supporting the re-ChIP analysis showing that they are present simultaneously in the regulatory region of *DYX1C1*. Already in Study I we showed that the polymorphism rs3743205 (-3G/A) affects the binding of TFII-I in the EMSA and expression levels in the luciferase assays. Therefore we studied if this polymorphism could affect the luciferase activity of the plasmid containing 132 bp long regulatory region of DYX1C1 (DYX132bp). The -3A variant showed a significant increase in the luciferase expression compared to the -3G variant, which is concordant with the previous luciferase results from Study I. Interestingly, E2 treatment had no apparent effect on the luciferase expression with the DYX132bp-3A plasmid. This again indicates that TFII-I binding is important for E2 mediated transcription of *DYX1C1*, since TFII-I binding to the -3A variant is reduced compared to the -3G variant.

Transcriptional regulation is a complex phenomenon where many factors, such as timing, cell type, the regulatory proteins and additional cellular factors present, play a role. The fact that not only the most proximal regulatory regions such as promoter, 5' and 3' UTRs, control the gene expression but also distant enhancer and repressor regions can affect the regulation (Sakabe and Nobrega 2010). In this thesis, we have mainly studied the genomic region around the rs3743205 near the translation initiation site of DYX1C1. This polymorphism was identified in the original paper by Taipale and co-workers when the A allele was indicated as the risk allele (Taipale *et al.* 2003). Many of the replication studies have showed the opposite, G as the risk allele and therefore interpreted their results as negative or not supporting the causal role of the SNP (Scerri *et al.* 2004; Wigg *et al.* 2004). Association to opposite alleles is not an uncommon phenomenon in complex diseases where environmental and interactions between different factors contribute to the susceptibility (Lin *et al.* 2007). Another level of complexity is added by the fact that the switch from G to A in this position leads to a

loss of CpG nucleotide since it is predicted to be situated within a CpG island. Thus, DNA methylation may play a role in *DYX1C1* regulation. This possibility should be further studied. In conclusion, we showed that all three SNPs can have an effect on the *DYX1C1* gene expression and therefore could explain the association with dyslexia.

Transcriptional regulation as a functional mechanism of the increased risk for dyslexia has also been suggested for the other dyslexia candidate genes such as *ROBO1*, *KIAA0319* and *DCDC2* (Hannula-Jouppi *et al.* 2005; Dennis *et al.* 2009; Meng *et al.* 2011). Indeed, this is a very intriguing explanation, since all these genes appear to have a modest effect on the disease susceptibility and since only a slight change in the expression levels during development can cause impairments in memory and learning related tasks as shown by the rodent studies (Threlkeld *et al.* 2007).

5.1.1 Transcription factors binding to the *DYX1C1* regulatory region

Because, we are studying a dyslexia susceptibility gene, the factors that regulate this gene should be expressed in brain, preferably during development and have functions in brain related tasks such as cognitive abilities. We have identified six different proteins (TFII-I, PARP1, SFPQ, c-MYB, RFX-1 and ERβ) that bind alone or as a complex to the proximal regulatory region of *DYX1C1* in SH-SY5Y cells. In addition, SP1 factor family proteins were also shown to bind to the distant regulatory regions of DYX1C1 in Study I. Here, I discuss why these factors are interesting in the context of *DYX1C1* regulation.

The *GTF2I* gene, encoding the TFII-I protein, is located on chromosome 7q11.23 region. TFII-I was initially characterized as a basal transcription factor binding to initiator sites of various promoters (Roy *et al.* 1991). In mouse, *Gtf2i* is expressed during early development and knockout of the gene is embryonic lethal (Enkhmandakh *et al.* 2009). Furthermore, exencephaly (open anterior neural tube) was found in the $Gtf2i^{-/-}$ and $Gtf2i^{+/-}$ mice, indicating that TFII-I has a role in early brain development. Interestingly, deletions of the genomic region where *GTF2I* is located are associated with the neurodevelopmental Williams-Beuren syndrome (WBS) in humans (Perez Jurado *et al.* 1998). Reading deficits have been described as one of the disabilities in WBS although the patients have mental retardation and therefore could not be given a proper diagnosis of dyslexia. TFII-I has been also connected to the estrogen signaling pathway by interacting with ERα and downregulating estrogen responsive genes (Ogura *et al.* 2006). In this thesis we also connect TFII-I to ERβ (III), further strengthening the role of TFII-I in E2 dependent gene regulation. TFII-I is probably one of the main players in the regulation of *DYX1C1* and its role can be modified by different co-factors as well as the cellular environment as shown in Studies I and III.

The SFPQ (splicing factor, proline/glutamine rich) that encodes the polypyrimidine tract binding protein associated splicing factor (PSF) was first identified as an essential pre-mRNA splicing factor (Patton *et al.* 1993). It has since then been shown to have multiple functions such as transcriptional corepression (Liu *et al.* 2011). Interestingly, the expression of Sfpq/Psf in mouse brain was found to be developmentally regulated, and it was found to be expressed in cortical plate during the last gestational and early postnatal weeks (Chanas-Sacre *et al.* 1999). This suggests that SPQ has role in neuronal

differentiation and establishing synaptic connections. Furthermore, Spfq was shown to be crucial for the normal brain development in zebrafish (Lowery *et al.* 2007). The *SFPQ* gene is located in a region that has been linked to SSD, language impairment, and dyslexia on chromosome 1p34-p36 (Table 2). SFPQ was seen to bind together with TFII-I and PARP1 to the regulatory region of DYX1C1 thus working as a co-regulator. However, the exact role of SFPQ in the regulation of *DYX1C1* remains to be determined.

PARP1 (Poly[ADP-ribose] polymerase 1) has been implicated in the regulation of gene expression through modification of TF binding (modifying histones to alter chromatin structure) or functioning as part of promoter binding complexes together with other DNA binding factors and co-activators (Kraus and Lis 2003). PARP1 also play an important role in the neuronal cell death and apoptosis and its activation is involved in a number of pathological conditions, including Parkinson's and Alzheimer's disease (Kauppinen and Swanson 2007). Moreover, PARP1 was shown to be involved in learning and long-term neuronal plasticity in the sea slug Aplysia californica and mice (Cohen-Armon *et al.* 2004; Hernandez *et al.* 2009; Fontan-Lozano *et al.* 2010). Similar to SFPQ, no direct binding of PARP1 has been demonstrated. Therefore PARP1 probably acts as a co-regulator modifying the effect of TFs (such as TFII-I) that bind directly to the regulatory regions of *DYX1C1*.

The regulatory factor 1 (RFX1), a prototype of the RFX family, has a conserved DNA binding domain of 76 amino acids that can bind to the X-box motif. RFX1 is highly expressed in the mammalian brain, compared to many other tissues and organs (Aftab *et al.* 2008). The function of RFX1 is unclear, however the knockout studies of *RFX* homologues in model organisms*, C. elegans* and *Drosophila*, indicated that the RFX has a role in neurons, especially in the differentiation of ciliated sensory neuron (Swoboda *et al.* 2000; Dubruille *et al.* 2002). A complete knockout of Rfx1 in mice leads to embryonic death (Feng *et al.* 2009). In rats, Rfx1 protein was shown to be expressed in the neurons and contribute to the regulation of the expression of the neuronal glutamate transporter type 3 (Ma *et al.* 2006). The fact that RFX1, at least in model organisms, has been connected to ciliogenesis and regulation of ciliary genes is very interesting. Cilia and ciliary genes have recently been connected to many human diseases (Hildebrandt *et al.* 2011) and we have implicated that the dyslexia candidate gene *DCDC2* has a function in cilia (Hildebrandt *et al.* 2011; Massinen *et al.* 2011). We have shown that RFX1 binds to the regulatory region of *DYX1C1* and could therefore modify the expression of *DYX1C1*. However, future research will show if the RFX factors and cilia play a role in the regulation of the dyslexia genes and in the etiology of dyslexia.

The c-MYB gene encodes a proto-oncogene transcription factor that has been mainly connected to human leukemia and developing hematopoietic system (Ramsay and Gonda 2008). It has also been shown to maintain cells in an undifferentiated state by repressing terminal differentiation and driving proliferation. In addition, c-MYB has an important role in neural progenitor cell proliferation and it is expressed in the subventricular zone during corticogenesis and later in adult brain (Malaterre *et al.* 2008). We could demonstrate a direct binding of c-MYB to the sequence nearby the DYX1C1 translation initiation site, but no effect of the c-MYB knockdown was seen in

the DYX1C1 expression levels. However, c-MYB could have a role in cell types different from the ones we have studied.

Specific protein-1 (SP1) belongs to the highly conserved Specificity Protein/Krüppel-like Factor (SP/KLF) transcription factor family and it is expressed in all mammalian cells regulating multiple genes (Li and Davie 2010). It has been suggested that SP1 might contribute to the complex etiology of schizophrenia by regulating several genes associated with the disorder (Ben-Shachar 2009). We have not studied the role of SP1 in the regulation of *DYX1C1* further; hence it should also be taken into account when looking at the complex picture of the regulation of *DYX1C1*. The ERβ will be discussed separately in the next chapter.

In conclusion, all the factors described above are needed for normal brain development, and therefore serve as highly interesting candidates in the regulation of *DYX1C1* and potentially for the other dyslexia candidate genes as well. However, all our studies have been done in cell lines, mainly in the SH-SY5Y cells; they might not represent the regulation in *in vivo* conditions. Therefore, it will be very important to replicate our findings in different systems. Furthermore, the participation of these factors in the etiology of dyslexia and more specifically in the regulation of other dyslexia susceptibility genes should be further studied.

5.2 DYX1C1 AND ESTROGEN RECEPTOR SIGNALING (II AND III)

Within cells, the molecules such as proteins work together e.g. in networks and signaling pathways to carry out biological functions. Therefore, to truly understand the function of a particular protein examination of its cellular role is needed. This will also help in identifying the pathways involved in different diseases. This chapter and also in 5.4 and 5.6, describes our results from experiments that we have done to understand the function of DYX1C1. In Study II, we investigated the interactions and involvement of DYX1C1 in the estrogen and ubiquitination pathways. Our hypothesis was based on previous publications of the interaction between DYX1C1 and U-box family ubiquitin protein ligase CHIP (Hatakeyama *et al.* 2004) and that CHIP has been shown to regulate the ERs (Tateishi *et al.* 2004; Fan *et al.* 2005; Tateishi *et al.* 2006). Estrogen signaling is a pathway where estrogen activates the ERs to modulate the cellular state at least in two ways, through the genomic and non-genomic pathways (*2.6*).

Indeed, we showed that the exogenous DYX1C1 and the ERs are co-localized in a specific pattern after E2 treatment in cells and that the proteins indeed interact. Interestingly, we could detect the native protein complexes of rat Dyx1c1/ERs along the extensions of primary rat hippocampal neurons using isPLA, which implies that this complex is present in the brain during development.

To examine functionality of this interaction, we overexpressed DYX1C1 and detected the protein levels of the ERs in MCF-7 and SH-SY5Y cells. The overexpression of *DYX1C1* decreased ERα and ERβ protein levels in a dose-dependent manner. Consistent with the interaction studies the decrease was greater in the E2-treated samples. Furthermore, this regulation was blocked using the proteasome inhibitor MG132 indicating that proteasome activity is needed for the DYX1C1-mediated effect.

Using luciferase reporter assays, we showed that the DYX1C1-mediated ER downregulation significantly diminished ERE_2-mediated reporter gene expression. These results indicate that DYX1C1 can negatively modulate transcriptional ER signaling pathways, at least the genomic pathway. The non-genomic pathway was not studied in this thesis; however, the localization of the native protein complex in the extensions of the neurons suggests that DYX1C1 could also modify the rapid signaling of the ERs in the cytoplasm. This should be further studied since the non-genomic E2 signaling has been suggested to be involved in modulating plasticity of the brain and cognition (Raz *et al.* 2008).

In Study III, we explored the estrogen pathway involvement in the regulation of *DYX1C1* as described in the section 5.1. We propose an estrogen dependent loop regulation of DYX1C1 and ERβ expression in SH-SY5Y cells based on the results from the Studies II and III. Upon E2 treatment, ERβ binds to the regulatory region of *DYX1C1* together with TFII-I and other TFs and co-regulators, enhancing the *DYX1C1* expression (Figure 11). DYX1C1 overexpression in turn targets liganded ERβ for proteasomal degradation with possible involvement of CHIP (Figure 11).

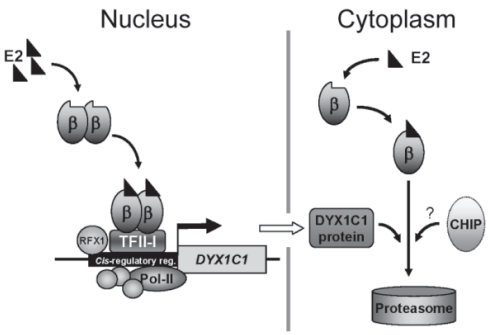

Figure 11 Schematic representation of the loop regulation between DYX1C1 and ERβ. In the presence of E2, ERβ is activated and recruited to the cis-regulatory region of DYX1C1, probably tethered to TFII-I. This augments *DYX1C1* mRNA expression. In turn, DYX1C1 (possibly together with CHIP) physically interacts with the liganded ERβ in the cytoplasm and targets it for proteasomal degradation.

Data presented in Studies II and III links DYX1C1 to the estrogen pathway. This is a very intriguing finding since there are many indications that estrogen signaling is important to brain development and cognitive functions. Most interestingly, studies on ERβ KO mice demonstrate deficits in the migration of post-mitotic neurons to the cortex (Wang *et al.* 2001; Wang *et al.* 2003), those defects are very similar to the phenotype produced by the *in utero* electroporation of shRNA against *DYX1C1* and to the anatomical findings in the post-mortem brains of dyslexic individuals (Galaburda *et al.* 1985; Wang *et al.* 2006). In the case of ERα, no brain abnormalities have been detected in the KO mouse, thus the importance of ERα in cortex development remains to be elucidated. However, it has been suggested that ERα is important for the function of CR cells, which are important for the correct positioning of cortical neurons (Gonzalez *et al.* 2007). In addition, it has been demonstrated that ERα is expressed in cells located in the VZ suggesting that it could have a role in neuronal proliferation (Martinez-Cerdeno *et al.* 2006).

Furthermore, *DYX1C1* may have a sex-specific risk effect on dyslexia which is interesting in the light of dyslexia being more common in boys than girls (Rutter *et al.* 2004; Dahdouh *et al.* 2009). A sex-specific effect was also seen in the translocation family, the boy with the translocation disrupting DYX1C1 had a more severe phenotype than the two girls (Nopola-Hemmi *et al.* 2000). I believe that DYX1C1 together with the ER are interesting candidates for further studies of neuro-endocrine signaling and the development of human cognitive functions and learning disorders such as dyslexia.

5.3 THE EFFECT OF DYX1C1 TO GLOBAL GENE EXPRESSION (IV)

The identification of differentially expressed transcripts has been successfully applied to investigating the gene functions and the underlying molecular mechanisms of differentiation, development and disease states. This can be done in many ways, i.e., by directly studying the expression levels in the affected tissue from patients compared to tissues from healthy individuals or after modulating the levels of the gene of interest in a cell/animal model.

There is no indication that DYX1C1 is a transcription factor thus it does not directly regulate the mRNA levels of other genes. However, it could affect gene expression by binding to and modulating the effect of different TFs such as the ERs. Therefore, we wanted to analyze the global gene expression pattern of a cell line after perturbation of *DYX1C1* levels in Study IV. To do this, we used the SH-SY5Y^{DYX1C1} overexpressing the full length *DYX1C1* and cells contain *DYX1C1* knockdown by siRNA for 48 h (siDYX1C1) together with corresponding controls and measured the expression levels with the Illumina HT-12v4 expression arrays.

Overall, we detected 379 probes corresponding 357 genes in the SH-SY5Y^{DYX1C1} samples and 88 probes corresponding to 87 genes in the siDYX1C1 samples to be differentially expressed (adj. $p<0.01$, $B>1$). Table 4 shows the 20 most significant probes for the comparisons. In total, 30 genes including *PDGFRA*, *SNAP91*, *CUX2*, *GAL*, *IL11RA*, *OLFM1* and *PDS5A* were differentially expressed in both comparisons.

Table 4 Top 20 differentially expressed probes obtained from the expression arrays studies.

ProbeID	Gene Symbol	Location	Gene name	logFC	P.Value	adj.P.Val	B
SH-SY5Y^{DYX1C1} – control							
4490221	RELN	7q22	Reelin	4.31	4.56e-11	2.16e-06	12.65
60731	PEX14	1p36.22	Peroxisomal biogenesis factor 14	2.22	1.23e-10	2.91e-06	12.28
4830376	ASCL1	12q23.2	Achaete-scute complex homolog 1 (Drosophila)	-5.38	3.08e-10	4.85e-06	11.88
430050	BAMBI	10p12.3-p11.2	BMP and activin membrane-bound inhibitor homolog (Xenopus laevis)	3.95	5.44e-10	6.44e-06	11.61
2760148	IL11RA	9p13	Interleukin 11 receptor, alpha	-2.03	7.94e-10	6.49e-06	11.42
1410403	LRRN2	1q32.1	Leucine rich repeat neuronal 2	-3.40	8.23e-10	6.49e-06	11.40
3140523	DYX1C1	15q21.3	Dyslexia susceptibility 1 candidate 1	3.61	1.43e-09	9.68e-06	11.11
110100	CGNL1	15q21.3	Cingulin-like 1	2.80	2.10e-09	1.12e-05	10.89
1580576	ITM2A	Xq13.3-q21.2	Integral membrane protein 2A	2.41	2.13e-09	1.12e-05	10.88
5700095	ARHGAP36	Xq26.1	Rho GTPase activating protein 36	-2.91	2.65e-09	1.14e-05	10.75
5690181	CEP44	4q34	Centrosomal protein 44kDa	-1.64	2.66e-09	1.14e-05	10.75
3990224	TNFRSF25	1p36.2	Tumor necrosis factor receptor superfamily, member 25	-2.20	3.32e-09	1.31e-05	10.62
1410408	ARID5B	10q21.2	AT rich interactive domain 5B (MRF1-like)	2.66	5.74e-09	2.09e-05	10.27
7570647	SNAP91	6q14.2	Synaptosomal-associated protein, 91kDa homolog (mouse)	1.71	7.29e-09	2.47e-05	10.12
6420470	PLD5	1q43	phospholipase D family, member 5	3.05	8.07e-09	2.55e-05	10.05
7510537	SCO2	22q13.33	SCO cytochrome oxidase deficient homolog 2 (yeast)	-1.65	9.65e-09	2.80e-05	9.93
6040431	GRM8	7q31.3-q32.1	Glutamate receptor, metabotropic 8	2.38	1.00e-08	2.80e-05	9.91
4120431	TMEM100	17q22	Transmembrane protein 100	4.63	1.30e-08	3.41e-05	9.73
4920148	ALDH1A3	15q26.3	Aldehyde dehydrogenase 1 family, member A3	-2.62	1.61e-08	4.02e-05	9.58
6660382	CDKN1A	6p21.2	Cyclin-dependent kinase inhibitor 1A (p21, Cip1)	1.70	1.89e-08	4.26e-05	9.47
2810315	S100A6	1q21	S100 calcium binding protein A6	-1.44	1.89e-08	4.26e-05	9.47
siDYX1C1 – siControl							
2640369	EZR	6q25.3	Ezrin	-1.61	6.86e-10	3.25e-05	9.92
6760386	LOC283267	11p13	Long intergenic non-protein coding RNA 294	-1.68	1.52e-08	0.00028	8.59
2450093	VCP	9p13.3	Valosin containing protein	-2.86	1.75e-08	0.00028	8.51
6650053	MCM3	6p12	Minichromosome maintenance complex component 3	-0.97	4.51e-08	0.00053	7.98
1410403	LRRN2	1q32.1	Leucine rich repeat neuronal 2	-2.07	5.84e-08	0.00055	7.83
1050544	SCAMP1	5q14.1	Secretory carrier membrane protein 1	-1.05	1.42e-07	0.00085	7.26
610356	ZDHHC9	Xq26.1	Zinc finger, DHHC-type containing 9	-1.71	1.47e-07	0.00085	7.23
4830376	ASCL1	12q23.2	Achaete-scute complex homolog 1 (Drosophila)	-2.60	1.57e-07	0.00085	7.19
2100612	EZR	6q25.3	Ezrin	-1.24	1.61e-07	0.00085	7.17
6370538	WBSCR22	7q11.23	Williams Beuren syndrome chromosome region 22	-1.32	1.93e-07	0.00092	7.05
2750647	C20ORF27	20p13	Chromosome 20 open reading frame 27	-0.89	2.97e-07	0.0013	6.75
6450066	CUX2	12q24.12	Cut-like homeobox 2	-1.65	3.21e-07	0.0013	6.70
3870255	PPPDE1	1q44	PPPDE peptidase domain containing 1	-0.98	4.10e-07	0.0015	6.52
7380110	CDK4	12q14	Cyclin-dependent kinase 4	-0.95	4.38e-07	0.0015	6.47
110592	SUSD2	22q11-q12	Sushi domain containing 2	-2.69	5.02e-07	0.0015	6.37
60731	PEX14	1p36.22	Peroxisomal biogenesis factor 14	0.83	5.49e-07	0.0015	6.31
3140520	TMOD1	9q22.3	Tropomodulin 1	1.01	5.55e-07	0.0015	6.30
6270646	MYC	8q24.21	V-myc myelocytomatosis viral oncogene homolog (avian)	-1.13	6.01e-07	0.0015	6.24
670408	PPP6C	9q33.3	Protein phosphatase 6, catalytic subunit	-0.62	6.15e-07	0.0015	6.22
5340017	H19	7	H19 fetal liver mRNA	-1.60	7.02e-07	0.0017	6.12
670286	CDK6	7q21-q22	Cyclin-dependent kinase 6	-0.90	8.05e-07	0.0018	6.02

To extract relevant information from these gene lists, e.g., canonical pathways and biological processes, we analyzed the gene ontology (GO) and KEGG biological pathway enrichment among the significantly perturbed genes. The most significant enrichment of genes was found in "cellular component movement", "cell migration" and "nervous system development" terms, corresponding to the previous evidence of the function of DYX1C1 (Figure 12). A total of 18 genes were classified in the "cell migration" term and six of those genes (*TWIST1*, *RELN*, *PHOX2B*, *NRCAM*, *DCX* and *PXMP3I*) were also categorized in the "neuron migration" term. Enrichment was also found for many brain related terms such as "neuron differentiation", "axonogenesis", and "neuron projection". In the KEGG pathway analysis, two interesting pathways were overrepresented among the genes "cell cycle pathway" and "focal adhesion pathway". Both of these pathways are needed for proper neuronal/cell migration (Lock *et al.* 2008; Pramparo *et al.* 2011).

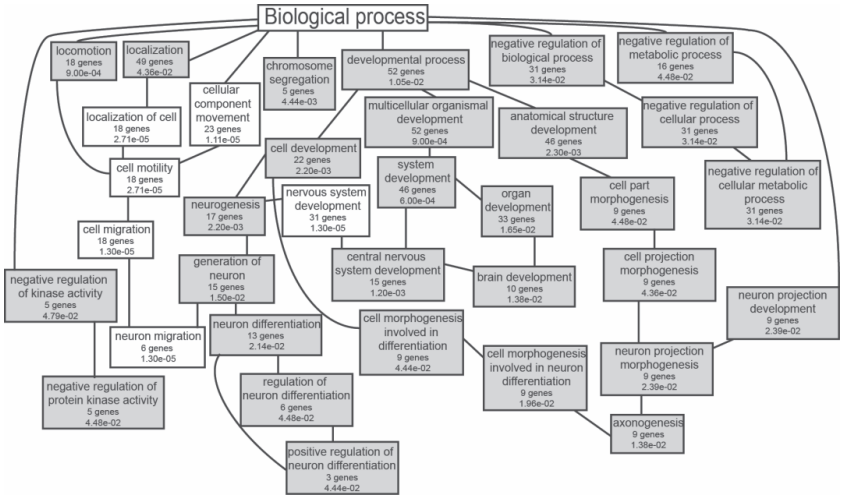

Figure 12 Gene ontology (biological process) enrichment among the upregulated genes in the SH-SY5Y^{DYX1C1} cell line.

Our gene expression profiling experiments after the perturbation of *DYX1C1* levels gives many interesting clues to understand the function of DYX1C1, not only in the biological processes as seen in the Figure 12 but also identified many genes that could serve as future candidates in the pathogenesis of dyslexia. Further analysis of the TFs regulating the affected genes would also serve as interesting follow up for these results since we might be able to identify the master regulators of the genes involved in these processes. I think the most intriguing finding is the fact that the two most studied lissencephaly genes are affected, namely *RELN* and *DCX.* As already mentioned in the introduction chapter *2.7*, Reelin is an important modulator of the neuronal migration pathway. It would be interesting to study if Reelin could be one factor causing the non-autonomous cell effects and the "over-migration" in the rat Dyx1c1 shRNA studies (Threlkeld *et al.* 2007; Currier *et al.* 2011). Reelin and other genes involved in the neuronal migration will be further discussed in the chapter *5.7*.

Recently, global gene expression analysis has also been used to study the effects of knockout of four highly relevant neuronal migration genes. This study showed that cell cycle as one of the commonly altered pathways (Pramparo *et al.* 2011). We could also detect it as an affected pathway in both knockdown and up-regulation experiments of DYX1C1 and also in the study of DCDC2 overexpression in primary hippocampal neurons (Massinen *et al.* 2011; Pramparo *et al.* 2011). Thus our results confirm and strengthen the fact that genes involved in the neuronal migration have a role in cell cycle.

In addition, differentiation and morphogenesis of neurons may be affected by DYX1C1. These have not been studied by us or others so far, however these are very interesting functions and show therefore be investigated. DYX1C1 appears to affect very similar biological processes as the transcription factor FOXP2 that regulates cognitive functions and language development, (Vernes *et al.* 2011).

Even though we were able to identify highly interesting genes and pathways affected by the perturbation of DYX1C1, we should take into account that the neuroblastoma cell line may not be the ideal model. Therefore these findings should be replicated in a different biological set up such as primary neurons or neuronal progenitor cells. One very interesting approach would be to use human induced pluripotent stem cell (iPSC) derived neurons which have already been predicted to be successful tools in the research of brain-related disorders (Dolmetsch and Geschwind 2011).

5.4 EVOLUTIONARY ANALYSIS OF DYX1C1 PROTEIN: IDENTIFICATION OF A NEW PROTEIN DOMAIN

By studying the nucleotide or protein sequence of a certain gene in different organisms, it is possible to extract relevant information about the functional domains in the protein and the conservation during evolution. Already in the original paper describing DYX1C1, its conservation was studied in other primates and mouse (Taipale *et al.* 2003). In Study IV, we conducted an extensive new search of Genbank using the blast server at NCBI to retrieve numerous sequences not only from vertebrates, but also from invertebrates, such as sea anemone and many single-celled eukaryotes. By sequence alignment, we could identify two additional conserved regions in the DYX1C1 amino acid sequence in addition to the TPRs and p23 domains described previously. The region after the p23 domain is conserved (DYX1C1 charged region, Figure 7), although to a lesser extent than the function protein domains. In front of the TPR repeats we identified a novel, highly conserved domain of 43 residues that we named "DYX1-domain". This domain was not detected in any other protein family by PSI-blast searches, so we concluded that it is a unique domain found only in the DYX1C1 protein. The function of the domain remains to be seen, but in our experiments it has very similar roles to the TPRs. It is needed for the migration phenotype produced by the overexpression of DYX1C1 (discussed more in 5.7) as well as for the centrosomal localization of the protein (discussed more in 5.6).

The sequence similarity of DYX1C1 in species is substantial; for example, the orthologous protein sequence from the choanoflagellate Monosiga brevis has a blast E-value of 3e-54 to the human DYX1C1 sequence, and is 34% identical over the whole

length. The protein domain structure is also very similar between species, especially the N-terminal part of the protein. The count of the TPRs varies in the lower species and it is actually lost in the Drosophila lineage. Surprisingly, each species or strain has only a single DYX1C1 ortholog which is particularly noteworthy for vertebrates, where many genes are duplicated. The fact that DYX1C1 is an ancient protein present in early eukaryotes suggests that it could be involved in basic cellular processes such as cell cycle or migration.

As said DYX1C1 is present in lower plants, however it has been lost in higher plants (e.g. Arabidopsis) and fungi, such as yeast. Curiously, it has also been lost in nematodes such as *C. elegans*. One interesting notion regarding this distribution: all the protists that have DYX1C1 are ciliated or flagellated while higher plants and yeast lack cilia. In the case of *C. elegans*, motile cilia have been lost. Intriguingly, this distribution is similar to Hedgehog-related proteins that are known to function in cilia (Burglin 2008). This correlation suggests that DYX1C1 protein may interact with a microtubule based machinery and perhaps associated with cilia.

5.5 PROTEIN INTERACTING PARTNERS FOR DYX1C1 (IV)

Recent studies have shown that susceptibility genes for many disease phenotypes often act together in the same biochemical pathway, multiprotein interaction networks and/or subcellular organelle (Vidal *et al.* 2011). There are many reasons why DYX1C1 protein would have multiple interacting partners; first the protein has two distinct domains, p23 and TPRs that are known to function in protein-protein interactions (ppi) and composing multiprotein complexes, and second the fact that older proteins tend to have more interactions than new ones. Therefore the identification of DYX1C1 interacting partners would provide important information of the biochemical pathways involved in addition to the global gene expression profiling discussed in 5.3.

Indeed, we and others have shown that DYX1C1 interacts with heat shock proteins, the CHIP protein, the ERs (II) and key proteins in the autophagy system (Hatakeyama *et al.* 2004; Behrends *et al.* 2010). In Study IV, we further analyzed the proteins associated with DYX1C1. First, we tested the possible interactions of DYX1C1 with LIS1, DCDC2 and KIAA0319. We chose those proteins based on their suggested involvement in dyslexia and/or neuronal migration. We detected an interaction between LIS1 and DYX1C1 (exo- and endogenous) using Co-IPs and isPLA. This interaction was mediated by the p23 domain of DYX1C1. We also detected an interaction between DCDC2 and DYX1C1, which was also mediated by the p23 domain. However, no interaction was detected between exogenous DYX1C1 and KIAA0319.

In addition, we wanted to screen for new potential interaction partners using Co-IPs combined with protein identification using liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS). By this means, we identified 66 new proteins associated with DYX1C1. The new protein association partners for DYX1C1 and also all the published interactions are presented in Figure 13. To complete the network of the DYX1C1 interaction partners, we conducted searches of the Pathways common database and found that 43 of the proteins in the DYX1C1 interactome have previously been shown to interact with each other (blue edges in Figure 13). The high connectivity

of the proteins suggested that DYX1C1 may be present in multiprotein complexes. In addition, we confirmed interaction for three proteins (TUBB2B, TUBA1C and Ataxin1) with additional Co-IPs followed by western blots.

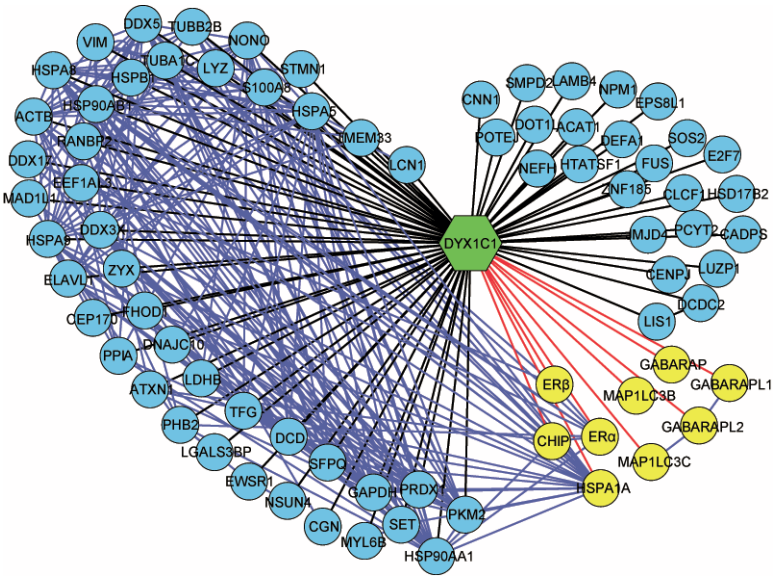

Figure 13 The DYX1C1 interactome. A schematic representation shows DYX1C1 interacting proteins identified in Study IV, databases and literature. Blue nodes represent interacting partners from the Study IV and yellow nodes from literature. Blue edges demonstrate the interactions between the proteins derived from the Pathways common database.

We also investigated the GO term enrichment among the DYX1C1 associated proteins. It was suggested that DYX1C1 could act as a co-chaperone for CHIP and heat shock proteins (Hatakeyama *et al.* 2004). This is in line with our GO analysis, where significant enrichment was found in "protein folding" and "response to unfolded protein" terms. Furthermore, "microtubule based process" and "microtubule organizing center organization" were also among the most significant biological processes. Similar results were seen in cellular component analysis with significant overrepresentation of "cytoskeletal proteins", more specifically "microtubule proteins". These results denote that DYX1C1 interacts with different cytoskeleton components needed for migration, such as microtubules and actin filaments. In addition, centrosomal proteins such as CEP170, CENPJ and NPM1 were enriched in the DYX1C1 supporting the localization of DYX1C1 to the centrosome. This pattern is very similar to other neuronal migration genes that promote the recruitment, stabilization and organization of microtubules and actin, which eventually drive neuronal migration and cell division. Similar to the global gene expression analysis results, by this approach we could identify genes that were crucial for the normal neuronal migration such as TUBB2B (Kumar *et al.* 2010; Uribe 2010). Interestingly, many transport related terms were among the significant terms such as "nucleocytoplasmic transport". This corresponds nicely with the dynamic localization pattern of DYX1C1 in cells discussed in 5.6.

Taken together the transcriptome results and the ppi results, we can conclude that DYX1C1 is indeed involved in the neuronal migration pathway affecting and interacting with multiple genes and their products in this pathway. However, it appears that DYX1C1 also has many other roles in the cells, such as functioning as co-chaperone, transporter and affecting cell cycle and cell death. More experiments are needed to dissect if these pathways are separate from each other or maybe these are all connected in some way. Our approaches have limitations and we may not have succeeded in identifying all the DYX1C1 interacting partners. In addition, some of the proteins from our screening could be false-positives. Similar to the transcriptome analysis, this analysis represent only a "snap-shot" of what is happening in the cells. If we would change the conditions of the cells for instance restrict the cell cycle, the interactome would look different. For this reason more screenings should be done in different cell lines and under different conditions. It would also be possible to increase the resolution by separating the cells in smaller fractions such as organelles and by this means provide more detailed information of the localization of the ppis.

5.6 LOCALIZATION OF THE DYX1C1 PROTEIN (I, II, III AND IV)

The localization of the DYX1C1 protein is very dynamic; it was determined to be nuclear when exogenous and endogenous expression was studied in COS-1 cells and cortical brain tissues in the first publication (Taipale *et al.* 2003). In contrast, DYX1C1 was also found in the cytoplasm when ischemic brain areas were studied. Later, Wang and co-workers indicated that exogenous DYX1C1 is concentrated to the cytoplasm in the COS-7 cells and neurons (Wang *et al.* 2006). They also determined the role of p23 and TPR domains in the DYX1C1 protein localization; DYX1C1 lacking the TPRs localized predominantly to the nucleus and DYX1C1 lacking the p23-domain localized predominantly to the cytoplasm (Wang *et al.* 2006). Since DYX1C1 appears to be localized differently in different cell types and tissues, we have studied the localization of DYX1C1 protein, both exogenous and endogenous, in the cell models used in our studies.

We demonstrated that the full length DYX1C1 protein upon overexpression was localized both to the nucleus and cytoplasm in SH-SY5Y (I), COS-7 (II), and MCF-7 (II) cells. When we examined the native DYX1C1 protein localization pattern in SH-SY5Y cells, we detected a predominantly cytoplasmic expression (III). We also studied the cellular localization of the DYX1C1 protein when the different domains were deleted. In COS-7 and MCF-7 cells, both DYX1C1Δp23 and DYX1C1ΔTPR were found in the nucleus and cytoplasm. In Study IV, we investigated the subcellular localization of the DYX1C1 protein. We showed that the exogenous DYX1C1 localized to the centrosome in SH-SY5Y cells. The TPR-domains were necessary for the localization to the centrosome and the deletion of the DYX1-domain weakened this localization.

We can conclude that the localization of DYX1C1 protein in cells and tissues is indeed very dynamic; one reason for this could be that it is part of transport system in the cells, where it needs to travel between the nucleus and cytoplasm. For instance, it has been shown that TPR-proteins are needed for the nucleocytoplasmic movement of the steroid hormones (Galigniana *et al.* 2010). Also the transport of signals/proteins in neurons is

crucial to maintain the development and function of the neuronal circuits. The localization of DYX1C1-ER protein complexes in the neurites could implicate involvement of DYX1C1 in the active transport of proteins in neurons (Ch'ng and Martin 2011).

Of course technical considerations should be taking into account when interpreting these localization results. Many of the experiments have been done with overexpression of DYX1C1 and may not represent the natural localization of the protein. For instance, the cells containing the overexpressing construct usually produce huge amount of protein that can lead to aggregation in unusual locations in the cells. Different antibodies can also produce different localization results depending on the affinity and specificity of the antibody. In Study III, we used a polyclonal antibody raised against the whole DYX1C1 protein, which should recognize all the splice variants, however it can also result in unspecific binding.

5.7 THE ROLE OF DYX1C1 IN CELL MIGRATION

Cell migration is a fundamental cellular process involving organization of cell body, plasma membrane and the cytoskeleton network. The cytoskeleton consists of three kinds of filaments: intermediate filaments, microtubules and actin filaments. In addition, accessory and motor proteins such as dyneins and kinesis are necessary for the cell movement and intracellular trafficking. Neuronal migration is one of the fundamental mechanisms together with axonal guidance underlying the organization of the brain (see also chapter 2.7).

Disturbed neuronal migration has been proposed as one of the underlying pathogenic pathways for dyslexia after inspections of the post-mortem brains of dyslexic individuals. This hypothesis was strengthened when RNAi studies against the rat Dyx1c1, Dcdc2 and Kiaa0319, showed that these genes are needed for normal neuronal migration during brain cortex development (Meng *et al.* 2005; Paracchini *et al.* 2006; Wang *et al.* 2006). In Study IV, we investigated if DYX1C1 could also influence the cell migration *in vitro*. Both knockdown and overexpression of DYX1C1 were used and the migration pattern of the SH-SY5Y cells was analyzed with live-cell imaging. Surprisingly, both conditions increased the velocity of random cell migration (Figure 14). Furthermore, we demonstrated that the C-terminal TPRs and the unique DYX1-domain are both needed for the increased migration (Figure 14B). This finding supported the results that impaired neuronal migration after silencing Dyx1c1 in rats could be rescued by the construct containing only the TPR-domains (Wang *et al.* 2006), and in addition implies the importance of the novel DYX1-domain identified in this thesis (5.4). A significant increase in the migration rate was also observed in the cells overexpressing DYX1C1Δp23 and two constructs containing dyslexia associated missense Val91Ile and nonsense Glu417XXX SNPs (rs317819126 and rs57809907, respectively), although the significance was less for the DYX1C1Δp23 and Val91Ile than for the full length protein (Figure 14B).

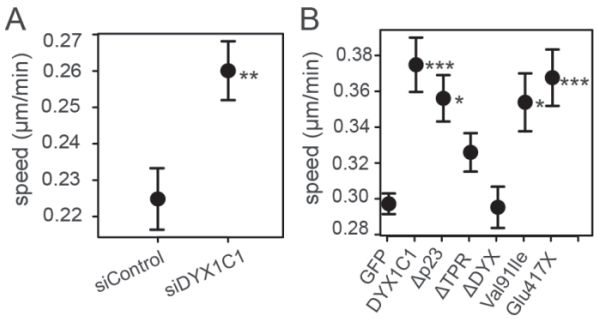

Figure 14 The velocity of random cell migration is affected by knockdown of DYX1C1 (A) and by the overexpression of different DYX1C1-GFP contructs (B) in SH-SY5Y cell line. Speed was calculated as mean migration rates (µm/min) ± SEM. Statistical significance was tested using was two-sided t-test (A) or ANOVA followed by Tukey's post hoc test (B). * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

By using live cell imaging we confirmed that DYX1C1 indeed affects cell migration. However, some differences are present when comparing our cell line migration results with the rat RNAi *in utero* electroporation studies; we demonstrated a clear effect of the overexpression of *DYX1C1*, in contrast to the rat study where it did not give any phenotype. Instead the knockdown showed a phenotype in both experiments. In rats, siRNA against *DYX1C1* restricted the neuronal migration near the SV zone and in adult rat brains a bimodal distribution was detected when about two-thirds of the neurons had migrated beyond their destination. This is in concordance with our results suggesting that when DYX1C1 levels are modulated the ability of the cell to control movement is affected.

In previous chapters I have provided evidence that DYX1C1 is involved in multiple molecular pathways, most of them connected to cell migration. Figure 15 shows a summary of these pathways, the connected cell organelles and structures. Selected examples of proteins and genes from the transcriptome and interactome experiments are placed in the pathways, but not all relationships between these and downstream signaling are presented in this summary figure. Overall, we can conclude that DYX1C1 has multiple roles in the cell and in different pathways. The next step will be to understand the timing and consequence of all these interactions and regulations. To take a step further, the genes in these pathways should be taken into consideration as possible functional candidates in the etiology of dyslexia. For instance, it would be useful to sequence genes such as LIS1, DCX and Reelin and their regulatory regions in dyslexic individuals.

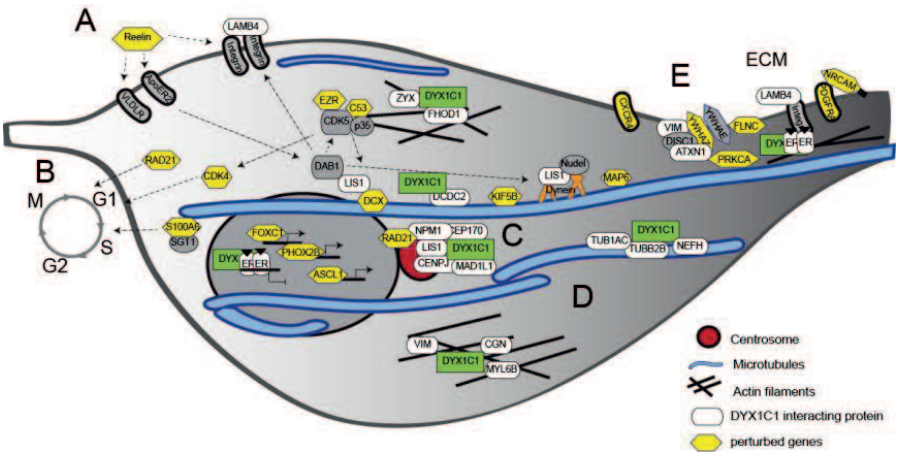

Figure 15 Pathways and cell organelles connected to cell migration and affected by DYX1C1**.** A) Reelin pathway B) Cell cycle pathway C) centrosome D) Cytoskeleton organization E) Focal adhesion (ECM=extracellular matrix). Arrows with dashed lines represent interactions or downstream signaling activation.

6 CONCLUDING REMARKS AND FUTURE PERSPECTIVES

To date many genes have been implicated as susceptibility factors for dyslexia. Studies of post-mortem dyslexic brains and knockdown of the dyslexia susceptibility genes in rats have provided evidence that subtle disturbances in neuronal migration could be the underlying biological cause of dyslexia. To take a step further we have focused on discovering regulatory and molecular networks of the first identified dyslexia gene *DYX1C1*. As a result of this thesis, we have connected DYX1C1 to several interesting molecular pathways, and can strengthen and further develop some of the already existing hypothesis of the biological causes of dyslexia. First, in this thesis we provide evidence that the estrogen pathway could be involved in the etiology of dyslexia. This supports the hypothesis that hormones are involved in the development of reading ability. Second, we found further support the involvement of the neuronal migration pathway in the etiology of dyslexia. It is highly interesting that we can connect DYX1C1 to the core genes, such as Reelin and LIS1, in this pathway. These and other genes that are involved in neuronal migration have been associated with many neurodevelopmental and psychiatric disorders. When analyzing results from GWAS, CNV and whole-exome studies for these different disorders (such as autism), it appears that the associated variations hit few distinct pathways such as neuronal migration and synapse development. It is possible that in the end all these phenotypes that we think are very distinct from each other actually fall into a same group of neuronal migration disorders. The fact that many of these conditions are also comorbid speaks in favor of this hypothesis.

In addition to the neuronal migration pathway, we have provided some evidence that the dyslexia candidate genes are connected to the centrosome-cilia complex (CCC). We recently published a study describing the involvement of DCDC2 in the regulation of cilia length and signaling in hippocampal neurons (Massinen *et al.* 2011). In addition, when looking at the regulation and function of DYX1C1 from many aspects, we could speculate that it may also have a role connected to the CCC. The centrosome localization of DYX1C1 provides evidence supporting this hypothesis. The cilium has been connected to many phenotypes in recent years, such as to Joubert's and Meckel's syndrome (Hildebrandt *et al.* 2011). Both of these syndromes have mental retardation as a feature. Thus, CCC could have a role in developing higher cognitive functions and therefore it could be also involved in development of reading ability. For now this is only speculation and future research will show if there is a connection between the CCC and the etiology of dyslexia.

Besides identifying the molecular pathways of genes associated to complex disorders, we should put more efforts in understanding the regulation of these genes. As we have shown with our results on DYX1C1, many factors and variations in the regulatory regions can affect gene expression. These small changes in gene expression are likely to be the cause of mild phenotypes such as dyslexia. Since many of the susceptibility genes are involved in important cellular processes, mutations disrupting the coding could result in more severe phenotypes such as mental retardation.

When looking back to year 2007 when I started as a PhD student in this project, all of the best dyslexia susceptibility genes such as DYX1C1, DCDC2 and KIAA0319 were already identified. However, everyone in the field was expecting for new gene discoveries from the ongoing GWA studies such as one from the Neurodys (www.neurodys.com). Unfortunately, no successful GWA study has been published for dyslexia, showing again how complex the phenotype is and how hard is to find the genetic factors underlying this disorder. Therefore, new interdisciplinary efforts using different strategies and approaches should be started to discover all the genetic components in dyslexia.

Interesting studies using different approaches to identify the genetic factors underlying dyslexia have already being published. For instance, when a speed perception, endophenotype of dyslexia measured by the mismatch negativity, was used as trait in a GWA study a genome-wide significance was found for several SNPs (Roeske *et al.* 2011). This suggests that it could be more beneficial to use different quantitative components of dyslexia as phenotypes in the GWA studies. Another approach would be to use the measures in brain structures and activation patterns as the phenotype. There is already evidence showing that variations in DCDC2 affect cortical morphology (Meda *et al.* 2008). Recently, an intronic SNP in the *FOXP2* gene was shown to affect the activation pattern in the left hemispheric areas connected to language and reading in dyslexic children (Wilcke *et al.* 2011). Identification of rare genetic variants in families has been a successful story in autism genetics; this might also be the case in some dyslexic families. Thus more research should be done to identify these variations.

The understanding of consequences of sequence variations, the gene-gene and gene-environmental interactions and other underlying genetic mechanisms together with the biological pathways are crucial when studying genetics of complex disorders. Even though this understanding has increased enormously, we still need to invent better tools for analyzing these mechanisms and connections between them. The identification of genetic causes of diseases is still very SNP and gene-centric, instead we should start thinking about pathways as biological units affected by the genomic variations and how these pathways are modulated upon environmental stimuli.

I have been asked many times during these years why are we studying dyslexia? And some of people still think that there is no such a phenotype as dyslexia. This always surprises me since dyslexia is the most common childhood learning disability affecting many children throughout the world. For a dyslexic child to have same opportunities in life as a child with normal reading skills, a proper diagnosis followed by remediation is needed. The progress of remedial modalities is based on a better understanding of the underlying biological and social processes. Therefore, current research in dyslexia, such as ours, aims at understanding the biological basis of this disorder. Hopefully, these findings can later guide novel hypotheses about how dyslexia is best diagnosed and optimal remedy planned. In addition, research on the biological basis of dyslexia may help us to better understand the origin, evolution, and development of high cognitive functions in humans.

7 ACKNOWLEDGEMENTS

The work presented in this thesis has been carried out at the Department of Biosciences and Nutrition, Karolinska Institutet during the years 2007 to 2011. I have been privileged to have such an interesting project, but most of all having so many people that have contributed, helped and supported me during these years. You have made this to a fantastic experience! Therefore I would like to express my deepest gratitude to all the people who have been part of my life during these years. I would especially like to thank:

My inspiring supervisor **Juha.** First of all, for answering to my first email almost eight years ago, accepting me as a summer student and convincing me to come to Stockholm instead of Helsinki! During these years you have always been supportive and enthusiastic about my work and carrier. I admire your talent, optimism and our passion to science! You have always listened to my ideas and given the opportunity to test those. This has given me the courage to become more and more independent as a scientist. I have also felt that I can come to you with other problems than scientific and discuss life in general, thank you for all those talks!

My fantastic co-supervisor **Isabel**, I don't know what I would have done without you! Thank you for letting me work independently, but always being there to give advices, discuss and listen to my worries especially when the westerns are not working! It is amazing to see that the work you started with just one clone for DYX1C1, and that we together have continued during these years, has led to a whole book of results! Thank you also for all the nice conference trips, and most of the all the everyday discussions and support in science and life!

Completing this thesis would not have been possible without our excellent dyslexia-team! It has been a privilege to work with so many wonderful and talented people. Thank you! **Myriam**, you introduced me to the project together with Heidi during my first summer in the group. Working with you was one of the main reasons why I wanted to come back. You mean a lot to me as a supervisor, colleague and most of all as a friend! **Hans**, for you excellent ideas and new methods that always gives an extra spice to our work! **Ingegerd**, for teaching me many things in the lab and your help with many of the projects! I don't know what the group would do without you! **Heidi**, for your friendship and support! **Gustaf,** for your ideas and expertise in IHC! Good luck with the studies and the PhD project! **Marco** for contributing so much to the dyslexia project, we are missing your analyzing expertise! **Satu,** for always having such excellent ideas! I really enjoyed our collaboration. **Marie-Estelle,** for providing us neurons to play with and great discussions about science and life! **Emma**, good luck with your project! **Vanina**, for your contributions to the dyslexia-project, best of luck in the future! **Eero**, for giving excellent comments and suggestions during our joined dyslexia meetings! **Gayathri**, good luck with the fishing experiments, I know we will have a lot of nice results from you! **Ettore**, welcome to the team and good luck!
I have also been honored to supervise many students during these years, it has been fun and I have learned a lot from you, thanks **Sini, Maria, Lisa** and **Kerttu**. Special thank

you goes to **Rachel** for helping me during this spring with projects in the lab and organizing my references!

In addition to our members of the dyslexia team, many collaborators have contributed to this thesis, especially thank you: **Jan-Åke Gustafsson** for help and many valuable comments and suggestions in the ER projects. **Eckardt**, your expertise made a huge difference to the ER project! **Ivan**, for taking your time to expand the ER project, I have really enjoyed working (emailing ☺) with you! **Sylvie**, for sharing your expertise in microscopes and having the patience with our long collaboration. Also for always being so positive and smiling! **Staffan Strömblad** for the initial suggestion of collaboration and helpful comments on the manuscript. **Morana**, for all the analysis you have done for me, good luck with your own project! **Thomas Börglin**, for your valuable analyses and giving more information about my favourite gene! **Tuula Nyman** for your expertise in mass spec and interactome studies! **Lisa**, for starting the fish project and also for reading and correcting this thesis! **Peter, Jan, Brian and Gilbert** for enjoyable collaboration and excellent ideas!

I would like to thank rest of the JKE group for wonderful working atmosphere, fun group trips and Christmas dinners! So thank you all the past and present members of the group: **Tiina** (thank you for all the nice discussions!), **Hong, Jingwen, Virpi, Elo, Nathalie, Cilla, Shintaro, Christina, Pu, Erja, Cecilia, Per** (thank you for the company in capoeira and nice parties!)**, Pauliina, Helena, Daniel, Marie, Mauro, Dario, Francesca A, Francesca B, Lovisa, Katja** and **Rasko.** Special thank you to all the PhD students in the group before me **Erik, Ulf, Heidi, Hanna, Anna, Sara and Christina** for showing me how things should be done! Also thanks to the MAF and the bioinformatics crews, **Kristina, Linda, Astrid, Aida, Päivi, Ville, Lotta, Malin, Tiina, Thomas**, **Rezin**… Special thanks to people in B.E.A for the microarray experiments especially **Marika** and **Fredrik**! Also Juha's group in Finland for fun group trips and collaborations, especially **Mari T.** and **Outi**!

Thank you for all the friends in Novum! Especially **Lotta, Anna E., Nina, Zoi, Karolina** and **Milica** for enjoyable company during lunches, coffee breaks and in the cell room! **Brigitte** for lunches, parties and fikas! Thank you goes also for all the administrative staff for always being friendly and helpful!

the TMK team: **Sara**, **Christina**, **Astrid**, **Linda, Hanna, Anna** and **Heidi**, for all the delicious food, wine and gossip! Thank you for all the fun, parties, sauna, trips… (this goes also to **Valtteri**, **Vilho**, **Tobbe**, **Kristian** and **Robert**)!It is really wonderful to have you as friends! **Anna**, tusen tack för korrigering av min svensk sammanfattning, jag skulle inte klarat utan dig ☺

Thank you all my lovely friends outside science or at least outside Novum:
Klara, for all the great time we have had together, already during the first summer in the "dagis"! It will be so much fun to have you, **Tyra** and **Petter** again here in Stockholm at least a while! **Judit**, for everything, all the nice fikas, dinners and shopping rounds, helping with the Swedish summary, the party and most of all your friendship! **Hannes,** for all the fun together. Soon we will celebrate your thesis!

My friends from capoeira in Finland, Stockholm and everywhere in the world, obrigada! Especially all the people at the **Academia de Gunga de Ouro**, for all the trainings, workshops and parties! Muito obrigada **Contramestre Boquinha** por tudo que você me ensinou na capoeira e também na vida!

Thank you for also my friends during the years in Oulu, especially **Inga** for all the fun together, **Antonio, Inka, Jari, Tiitu, Janne, Tero, Riikka, Anna**….! For all the people in the biology courses! Especially **Jenni**, it was so much fun to have you there with me in all the courses and also having fun outside the university! **Satu**, for all the visits to Stockholm, I hope I will have time soon to visit you!
Thanks **Päivi**, for great fun always, I miss going out dancing with you!

All the old friends from Rovaniemi, especially my girls, **Tea, Emma, Maria, Kaisa Monna** and **Emilia** and your wonderful "extensions" (meaning husbands, boyfriends and your babies!), thank you for all the great time that we have shared, all the laughs and tears❤ I miss you so much, and we should meet more often but I know you are always there for me!

And my family, I miss you all every day! **Äiti**, thank you for all the care and support that you have given me and for never questioning by decisions ☺ **Isi**, thank you for all the support and saying that I can do whatever I want in life! **Roope** and **Rami** you are the best young brothers ever! **Tintti** for all the nice time in Suhosenmutka! **Vinski**, hope you are having time of your life in Australia and Asia! **Tanja**, my dear sister, there are so many things that I want to thank you! for supporting and helping with all possible things, all the fun trips to abroad…also for correcting my Finnish, it is really needed! Thank you and **Teemu** for making the most perfect nephew **Niilo**. I cannot wait to meet the next one! **Tarja** and **Emma** for making me feel like home in Stockholm when I first came here and all the dinners and support! Aunt **Tarja** for all the nice dinners and place to stay in Helsinki and always remember me with greeting cards! Kiitos Eeva-mummo kaikesta tuesta! Kiitos myös kaikille muille sukulaisille! **Kirsti** ja **Veijo**, kiitos kun olette minut vastaan kuin oman tyttären. Rantsilaan on aina mukava tulla rentoutumaan ja nauttimaan maalaiselämästä! **Tiina, Eero, Hilla** and **Luukas**, for all the visits, hopefully we will meet even more often in the future!

and finally **Jaakko**, love of my life! You make me happy every day and I feel so lucky to have you in my life ❤ I want to thank you for all the big things such as moving to Stockholm with me but also for the all small everyday things like making me a cup of coffee every morning and helping with the cover figure! I cannot wait for our new adventures whatever they might be. Soon we will meet our own genetic experiment!

The work in this thesis was financially supported by Osk Huttunen Foundation, Swedish Royal Bank Tercentennial Foundation, Swedish Brain Foundation (Hjärnfonden) and European Union (CRESCENDO).

8 REFERENCES

Aftab, S., L. Semenec, J. S. Chu and N. Chen (2008). Identification and characterization of novel human tissue-specific RFX transcription factors. *BMC Evol Biol* **8**: 226.

Amet, L. E., S. E. Lauri, A. Hienola*, et al.* (2001). Enhanced hippocampal long-term potentiation in mice lacking heparin-binding growth-associated molecule. *Mol Cell Neurosci* **17**(6): 1014-1024.

Andrews, W., M. Barber, L. R. Hernadez-Miranda*, et al.* (2008). The role of Slit-Robo signaling in the generation, migration and morphological differentiation of cortical interneurons. *Dev Biol* **313**(2): 648-658.

Andrews, W., A. Liapi, C. Plachez*, et al.* (2006). Robo1 regulates the development of major axon tracts and interneuron migration in the forebrain. *Development* **133**(11): 2243-2252.

Anthoni, H., L. E. Sucheston, B. A. Lewis*, et al.* (2011).The Aromatase Gene CYP19A1: Several Genetic and Functional Lines of Evidence Supporting a Role in Reading, Speech and Language.*Manuscript*

Anthoni, H., M. Zucchelli, H. Matsson*, et al.* (2007). A locus on 2p12 containing the co-regulated MRPL19 and C2ORF3 genes is associated to dyslexia. *Hum Mol Genet* **16**(6): 667-677.

Assadi, A. H., G. Zhang, U. Beffert*, et al.* (2003). Interaction of reelin signaling and Lis1 in brain development. *Nat Genet* **35**(3): 270-276.

Bai, J., R. L. Ramos, J. B. Ackman*, et al.* (2003). RNAi reveals doublecortin is required for radial migration in rat neocortex. *Nat Neurosci* **6**(12): 1277-1283.

Bakwin, H. (1973). Reading disability in twins. *Dev Med Child Neurol* **15**(2): 184-187.

Bates, T. C., P. A. Lind, M. Luciano*, et al.* (2010). Dyslexia and DYX1C1: deficits in reading and spelling associated with a missense mutation. *Mol Psychiatry* **15**(12): 1190-1196.

Bates, T. C., M. Luciano, A. Castles*, et al.* (2007). Replication of reported linkages for dyslexia and spelling and suggestive evidence for novel regions on chromosomes 4 and 17. *Eur J Hum Genet* **15**(2): 194-203.

Bates, T. C., M. Luciano, S. E. Medland*, et al.* (2011). Genetic variance in a component of the language acquisition device: ROBO1 polymorphisms associated with phonological buffer deficits. *Behav Genet* **41**(1): 50-57.

Beech, J. R. and M. W. Beauvois (2006). Early experience of sex hormones as a predictor of reading, phonology, and auditory perception. *Brain Lang* **96**(1): 49-58.

Behrends, C., M. E. Sowa, S. P. Gygi and J. W. Harper (2010). Network organization of the human autophagy system. *Nature* **466**(7302): 68-76.

Bellini, G., C. Bravaccio, F. Calamoneri*, et al.* (2005). No evidence for association between dyslexia and DYX1C1 functional variants in a group of children and adolescents from Southern Italy. *J Mol Neurosci* **27**(3): 311-314.

Ben-Shachar, D. (2009). The interplay between mitochondrial complex I, dopamine and Sp1 in schizophrenia. *J Neural Transm* **116**(11): 1383-1396.

Berger, M., W. Yule and M. Rutter (1975). Attainment and adjustment in two geographical areas. II--The prevalence of specific reading retardation. *Br J Psychiatry* **126**: 510-519.

Best, M. and J. B. Demb (1999). Normal planum temporale asymmetry in dyslexics with a magnocellular pathway deficit. *Neuroreport* **10**(3): 607-612.

Biedler, J. L., S. Roffler-Tarlov, M. Schachner and L. S. Freedman (1978). Multiple neurotransmitter synthesis by human neuroblastoma cell lines and clones. *Cancer Res* **38**(11 Pt 1): 3751-3757.

Bielle, F., A. Griveau, N. Narboux-Neme*, et al.* (2005). Multiple origins of Cajal-Retzius cells at the borders of the developing pallium. *Nat Neurosci* **8**(8): 1002-1012.

Blatch, G. L. and M. Lassle (1999). The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. *Bioessays* **21**(11): 932-939.

Borecki, I. B. and B. K. Suarez (2001). Linkage and association: basic concepts. *Adv Genet* **42**: 45-66.

Brandon, N. J., J. K. Millar, C. Korth*, et al.* (2009). Understanding the role of DISC1 in psychiatric disease and during normal development. *J Neurosci* **29**(41): 12768-12775.

Brewer, G. J. and C. W. Cotman (1989). Survival and growth of hippocampal neurons in defined medium at low density: advantages of a sandwich culture technique or low oxygen. *Brain Res* **494**(1): 65-74.

Brkanac, Z., N. H. Chapman, M. M. Matsushita, *et al.* (2007). Evaluation of candidate genes for DYX1 and DYX2 in families with dyslexia. *Am J Med Genet B Neuropsychiatr Genet* **144B**(4): 556-560.

Bugge, M., G. Bruun-Petersen, K. Brondum-Nielsen, *et al.* (2000). Disease associated balanced chromosome rearrangements: a resource for large scale genotype-phenotype delineation in man. *J Med Genet* **37**(11): 858-865.

Buonincontri, R., I. Bache, A. Silahtaroglu, *et al.* (2011). A cohort of balanced reciprocal translocations associated with dyslexia: identification of two putative candidate genes at DYX1. *Behav Genet* **41**(1): 125-133.

Burbridge, T. J., Y. Wang, A. J. Volz, *et al.* (2008). Postnatal analysis of the effect of embryonic knockdown and overexpression of candidate dyslexia susceptibility gene homolog Dcdc2 in the rat. *Neuroscience* **152**(3): 723-733.

Burglin, T. R. (2008). The Hedgehog protein family. *Genome Biol* **9**(11): 241.

Cardon, L. R., S. D. Smith, D. W. Fulker, *et al.* (1994). Quantitative trait locus for reading disability on chromosome 6. *Science* **266**(5183): 276-279.

Carroll, S. B. (2008). Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* **134**(1): 25-36.

Casanova, M. F., J. Araque, J. Giedd and J. M. Rumsey (2004). Reduced brain size and gyrification in the brains of dyslexic patients. *J Child Neurol* **19**(4): 275-281.

Castles, A. and M. Coltheart (1993). Varieties of developmental dyslexia. *Cognition* **47**(2): 149-180.

Ch'ng, T. H. and K. C. Martin (2011). Synapse-to-nucleus signaling. *Curr Opin Neurobiol* **21**(2): 345-352.

Chahwan, R., S. N. Wontakal and S. Roa (2011). The multidimensional nature of epigenetic information and its role in disease. *Discov Med* **11**(58): 233-243.

Chanas-Sacre, G., C. Mazy-Servais, R. Wattiez, *et al.* (1999). Identification of PSF, the polypyrimidine tract-binding protein-associated splicing factor, as a developmentally regulated neuronal protein. *J Neurosci Res* **57**(1): 62-73.

Chang, B. S., T. Katzir, T. Liu, *et al.* (2007). A structural basis for reading fluency: white matter defects in a genetic brain malformation. *Neurology* **69**(23): 2146-2154.

Chang, B. S., J. Ly, B. Appignani, *et al.* (2005). Reading impairment in the neuronal migration disorder of periventricular nodular heterotopia. *Neurology* **64**(5): 799-803.

Chapman, N. H., R. P. Igo, J. B. Thomson, *et al.* (2004). Linkage analyses of four regions previously implicated in dyslexia: confirmation of a locus on chromosome 15q. *Am J Med Genet B Neuropsychiatr Genet* **131B**(1): 67-75.

Chen, Y., M. Zhao, S. Wang, *et al.* (2009). A novel role for DYX1C1, a chaperone protein for both Hsp70 and Hsp90, in breast cancer. *J Cancer Res Clin Oncol* **135**(9): 1265-1276.

Cirulli, E. T. and D. B. Goldstein (2010). Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat Rev Genet* **11**(6): 415-425.

Cohen-Armon, M., L. Visochek, A. Katzoff, *et al.* (2004). Long-term memory requires polyADP-ribosylation. *Science* **304**(5678): 1820-1822.

Conrad, D. F., D. Pinto, R. Redon, *et al.* (2010). Origins and functional impact of copy number variation in the human genome. *Nature* **464**(7289): 704-712.

Consortium, G. P. (2010). A map of human genome variation from population-scale sequencing. *Nature* **467**(7319): 1061-1073.

Consortium, I. H. (2005). A haplotype map of the human genome. *Nature* **437**(7063): 1299-1320.

Consortium, W. T. C. C. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**(7145): 661-678.

Cookson, W., L. Liang, G. Abecasis, *et al.* (2009). Mapping complex disease traits with global gene expression. *Nat Rev Genet* **10**(3): 184-194.

Cooper, J. A. (2008). A mechanism for inside-out lamination in the neocortex. *Trends Neurosci* **31**(3): 113-119.

Cope, N., D. Harold, G. Hill, *et al.* (2005). Strong evidence that KIAA0319 on chromosome 6p is a susceptibility gene for developmental dyslexia. *Am J Hum Genet* **76**(4): 581-591.

Cope, N. A., G. Hill, M. van den Bree, *et al.* (2005). No support for association between dyslexia susceptibility 1 candidate 1 and developmental dyslexia. *Mol Psychiatry* **10**(3): 237-238.

Couto, J. M., L. Gomez, K. Wigg, *et al.* (2008). The KIAA0319-like (KIAA0319L) gene on chromosome 1p34 as a candidate for reading disabilities. *J Neurogenet* **22**(4): 295-313.

Couto, J. M., I. Livne-Bar, K. Huang, *et al.* (2010). Association of reading disabilities with regions marked by acetylated H3 histones in KIAA0319. *Am J Med Genet B Neuropsychiatr Genet* **153B**(2): 447-462.

Currier, T. A., M. A. Etchegaray, J. L. Haight, *et al.* (2011). The effects of embryonic knockdown of the candidate dyslexia susceptibility gene homologue Dyx1c1 on the distribution of GABAergic neurons in the cerebral cortex. *Neuroscience* **172**: 535-546.

D'Arcangelo, G., R. Homayouni, L. Keshvara, *et al.* (1999). Reelin is a ligand for lipoprotein receptors. *Neuron* **24**(2): 471-479.

D'Arcangelo, G., G. G. Miao, S. C. Chen, *et al.* (1995). A protein related to extracellular matrix proteins deleted in the mouse mutant reeler. *Nature* **374**(6524): 719-723.

Dahdouh, F., H. Anthoni, I. Tapia-Paez, *et al.* (2009). Further evidence for DYX1C1 as a susceptibility factor for dyslexia. *Psychiatr Genet* **19**(2): 59-63.

Dahlman-Wright, K., V. Cavailles, S. A. Fuqua, *et al.* (2006). International Union of Pharmacology. LXIV. Estrogen receptors. *Pharmacol Rev* **58**(4): 773-781.

de Kovel, C. G., F. A. Hol, J. G. Heister, *et al.* (2004). Genomewide scan identifies susceptibility locus for dyslexia on Xq27 in an extended Dutch family. *J Med Genet* **41**(9): 652-657.

Deffenbacher, K. E., J. B. Kenyon, D. M. Hoover, *et al.* (2004). Refinement of the 6p21.3 quantitative trait locus influencing dyslexia: linkage and association analyses. *Hum Genet* **115**(2): 128-138.

DeFries, J. C., D. W. Fulker and M. C. LaBuda (1987). Evidence for a genetic aetiology in reading disability of twins. *Nature* **329**(6139): 537-539.

Dennis, M. Y., S. Paracchini, T. S. Scerri, *et al.* (2009). A common variant associated with dyslexia reduces expression of the KIAA0319 gene. *PLoS Genet* **5**(3): e1000436.

Deutsch, S. I., J. A. Burket and E. Katz (2010). Does subtle disturbance of neuronal migration contribute to schizophrenia and other neurodevelopmental disorders? Potential genetic mechanisms with possible treatment implications. *Eur Neuropsychopharmacol* **20**(5): 281-287.

Dimitrov, A., V. Paupe, C. Gueudry, *et al.* (2009). The gene responsible for Dyggve-Melchior-Clausen syndrome encodes a novel peripheral membrane protein dynamically associated with the Golgi apparatus. *Hum Mol Genet* **18**(3): 440-453.

Dolmetsch, R. and D. H. Geschwind (2011). The human brain in a dish: the promise of iPSC-derived neurons. *Cell* **145**(6): 831-834.

Drinjakovic, J., H. Jung, D. S. Campbell, *et al.* (2010). E3 ligase Nedd4 promotes axon branching by downregulating PTEN. *Neuron* **65**(3): 341-357.

Du, P., W. A. Kibbe and S. M. Lin (2008). lumi: a pipeline for processing Illumina microarray. *Bioinformatics* **24**(13): 1547-1548.

Duara, R., A. Kushch, K. Gross-Glenn, *et al.* (1991). Neuroanatomic differences between dyslexic and normal readers on magnetic resonance imaging scans. *Arch Neurol* **48**(4): 410-416.

Dubruille, R., A. Laurencon, C. Vandaele, *et al.* (2002). Drosophila regulatory factor X is necessary for ciliated sensory neuron differentiation. *Development* **129**(23): 5487-5498.

Duncan, D., N. Prodduturi and B. Zhang (2010). WebGestalt2: an updated and expanded version of the Web-based Gene Set Analysis Toolkit. *BMC Bioinformatics* **11**(Suppl 4): P10.

Eckert, M. A., C. M. Leonard, T. L. Richards, *et al.* (2003). Anatomical correlates of dyslexia: frontal and cerebellar findings. *Brain* **126**(Pt 2): 482-494.

Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**: 113.

Enkhmandakh, B., A. V. Makeyev, L. Erdenechimeg, *et al.* (2009). Essential functions of the Williams-Beuren syndrome-associated TFII-I genes in embryonic development. *Proc Natl Acad Sci U S A* **106**(1): 181-186.

Fagerheim, T., P. Raeymaekers, F. E. Tonnessen, *et al.* (1999). A new gene (DYX3) for dyslexia is located on chromosome 2. *J Med Genet* **36**(9): 664-669.

Fan, M., A. Park and K. P. Nephew (2005). CHIP (carboxyl terminus of Hsc70-interacting protein) promotes basal and geldanamycin-induced degradation of estrogen receptor-alpha. *Mol Endocrinol* **19**(12): 2901-2914.
Faulkner, G. J., Y. Kimura, C. O. Daub*, et al.* (2009). The regulated retrotransposon transcriptome of mammalian cells. *Nat Genet* **41**(5): 563-571.
Felker, M. V., L. M. Walker, D. K. Sokol*, et al.* (2011). Early cognitive and behavioral problems in children with nodular heterotopia. *Epilepsy Behav*.
Feng, C., W. Xu and Z. Zuo (2009). Knockout of the regulatory factor X1 gene leads to early embryonic lethality. *Biochem Biophys Res Commun* **386**(4): 715-717.
Fisher, S. E. and J. C. DeFries (2002). Developmental dyslexia: genetic dissection of a complex cognitive trait. *Nat Rev Neurosci* **3**(10): 767-780.
Fisher, S. E., C. Francks, A. J. Marlow*, et al.* (2002). Independent genome-wide scans identify a chromosome 18 quantitative-trait locus influencing dyslexia. *Nat Genet* **30**(1): 86-91.
Fisher, S. E., A. J. Marlow, J. Lamb*, et al.* (1999). A quantitative-trait locus on chromosome 6p influences different aspects of developmental dyslexia. *Am J Hum Genet* **64**(1): 146-156.
Fontan-Lozano, A., I. Suarez-Pereira, A. Horrillo*, et al.* (2010). Histone H1 poly[ADP]-ribosylation regulates the chromatin alterations required for learning consolidation. *J Neurosci* **30**(40): 13305-13313.
Francks, C., S. E. Fisher, I. L. MacPhie*, et al.* (2002). A genomewide linkage screen for relative hand skill in sibling pairs. *Am J Hum Genet* **70**(3): 800-805.
Francks, C., S. E. Fisher, R. K. Olson*, et al.* (2002). Fine mapping of the chromosome 2p12-16 dyslexia susceptibility locus: quantitative association analysis and positional candidate genes SEMA4F and OTX1. *Psychiatr Genet* **12**(1): 35-41.
Francks, C., S. Paracchini, S. D. Smith*, et al.* (2004). A 77-kilobase region of chromosome 6p22.2 is associated with dyslexia in families from the United Kingdom and from the United States. *Am J Hum Genet* **75**(6): 1046-1058.
Frazer, K. A., D. G. Ballinger, D. R. Cox*, et al.* (2007). A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**(7164): 851-861.
Fredriksson, S., M. Gullberg, J. Jarvius*, et al.* (2002). Protein detection using proximity-dependent DNA ligation assays. *Nat Biotechnol* **20**(5): 473-477.
Fried, G., E. Andersson, L. Csoregh*, et al.* (2004). Estrogen receptor beta is expressed in human embryonic brain cells and is regulated by 17beta-estradiol. *Eur J Neurosci* **20**(9): 2345-2354.
Friend, A., J. C. DeFries and R. K. Olson (2008). Parental education moderates genetic influences on reading disability. *Psychol Sci* **19**(11): 1124-1130.
Fugger, H. N., T. C. Foster, J. Gustafsson and E. F. Rissman (2000). Novel effects of estradiol and estrogen receptor alpha and beta on cognitive function. *Brain Res* **883**(2): 258-264.
Gabel, L. A., I. Marin, J. J. Loturco*, et al.* (2011). Mutation of the dyslexia-associated gene Dcdc2 impairs LTM and visuo-spatial performance in mice. *Genes Brain Behav*.
Galaburda, A. M. and T. L. Kemper (1978). Auditory cytoarchitectonic abnormalities in a case of familial developmental dyslexia. *Trans Am Neurol Assoc* **103**: 262-265.
Galaburda, A. M., G. F. Sherman, G. D. Rosen*, et al.* (1985). Developmental dyslexia: four consecutive patients with cortical anomalies. *Ann Neurol* **18**(2): 222-233.
Galigniana, M. D., P. C. Echeverria, A. G. Erlejman and G. Piwien-Pilipuk (2010). Role of molecular chaperones and TPR-domain proteins in the cytoplasmic transport of steroid receptors and their passage through the nuclear pore. *Nucleus* **1**(4): 299-308.
Garcia-Ranea, J. A., G. Mirey, J. Camonis and A. Valencia (2002). p23 and HSP20/alpha-crystallin proteins define a conserved sequence domain present in other eukaryotic protein families. *FEBS Lett* **529**(2-3): 162-167.
Gayan, J., S. D. Smith, S. S. Cherny*, et al.* (1999). Quantitative-trait locus for specific language and reading deficits on chromosome 6p. *Am J Hum Genet* **64**(1): 157-164.
Georgiewa, P., R. Rzanny, C. Gaser*, et al.* (2002). Phonological processing in dyslexic children: a study combining functional imaging and event related potentials. *Neurosci Lett* **318**(1): 5-8.
Gershon, E. S., N. Alliey-Rodriguez and C. Liu (2011). After GWAS: searching for genetic risk for schizophrenia and bipolar disorder. *Am J Psychiatry* **168**(3): 253-256.
Geschwind, N. (1965). Disconnexion syndromes in animals and man. I. *Brain* **88**(2): 237-294.

Geschwind, N. and A. M. Galaburda (1985a). Cerebral lateralization. Biological mechanisms, associations, and pathology: I. A hypothesis and a program for research. *Arch Neurol* **42**(5): 428-459.
Geschwind, N. and A. M. Galaburda (1985b). Cerebral lateralization. Biological mechanisms, associations, and pathology: II. A hypothesis and a program for research. *Arch Neurol* **42**(6): 521-552.
Gilger, J. W., B. F. Pennington and J. C. DeFries (1992). A twin study of the etiology of comorbidity: attention-deficit hyperactivity disorder and dyslexia. *J Am Acad Child Adolesc Psychiatry* **31**(2): 343-348.
Ginsburg, G. S. and H. F. Willard (2009). Genomic and personalized medicine: foundations and applications. *Transl Res* **154**(6): 277-287.
Glazier, A. M., J. H. Nadeau and T. J. Aitman (2002). Finding genes that underlie complex traits. *Science* **298**(5602): 2345-2349.
Gleeson, J. G., K. M. Allen, J. W. Fox*, et al.* (1998). Doublecortin, a brain-specific gene mutated in human X-linked lissencephaly and double cortex syndrome, encodes a putative signaling protein. *Cell* **92**(1): 63-72.
Gleeson, J. G., P. T. Lin, L. A. Flanagan and C. A. Walsh (1999). Doublecortin is a microtubule-associated protein and is expressed widely by migrating neurons. *Neuron* **23**(2): 257-271.
Goebl, M. and M. Yanagida (1991). The TPR snap helix: a novel protein repeat motif from mitosis to transcription. *Trends Biochem Sci* **16**(5): 173-177.
Gonzalez, M., A. Cabrera-Socorro, C. G. Perez-Garcia*, et al.* (2007). Distribution patterns of estrogen receptor alpha and beta in the human cortex and hippocampus during development and adulthood. *J Comp Neurol* **503**(6): 790-802.
Grigorenko, E. L. (2001). Developmental dyslexia: an update on genes, brains, and environments. *J Child Psychol Psychiatry* **42**(1): 91-125.
Grigorenko, E. L., F. B. Wood, M. S. Meyer*, et al.* (1997). Susceptibility loci for distinct components of developmental dyslexia on chromosomes 6 and 15. *Am J Hum Genet* **60**(1): 27-39.
Grigorenko, E. L., F. B. Wood, M. S. Meyer and D. L. Pauls (2000). Chromosome 6p influences on different dyslexia-related cognitive processes: further confirmation. *Am J Hum Genet* **66**(2): 715-723.
Grigorenko, E. L., F. B. Wood, M. S. Meyer*, et al.* (2001). Linkage studies suggest a possible locus for developmental dyslexia on chromosome 1p. *Am J Med Genet* **105**(1): 120-129.
Hallgren, B. (1950). Specific dyslexia (congenital word-blindness); a clinical and genetic study. *Acta Psychiatr Neurol Suppl* **65**: 1-287.
Hannula-Jouppi, K., N. Kaminen-Ahola, M. Taipale*, et al.* (2005). The axon guidance receptor gene ROBO1 is a candidate gene for developmental dyslexia. *PLoS Genet* **1**(4): e50.
Harlaar, N., F. M. Spinath, P. S. Dale and R. Plomin (2005). Genetic influences on early word recognition abilities and disabilities: a study of 7-year-old twins. *J Child Psychol Psychiatry* **46**(4): 373-384.
Harold, D., S. Paracchini, T. Scerri*, et al.* (2006). Further evidence that the KIAA0319 gene confers susceptibility to developmental dyslexia. *Mol Psychiatry* **11**(12): 1085-1091, 1061.
Hatakeyama, S., M. Matsumoto, M. Yada and K. I. Nakayama (2004). Interaction of U-box-type ubiquitin-protein ligases (E3s) with molecular chaperones. *Genes Cells* **9**(6): 533-548.
Hawke, J. L., S. J. Wadsworth and J. C. DeFries (2006). Genetic influences on reading difficulties in boys and girls: the Colorado twin study. *Dyslexia* **12**(1): 21-29.
Herculano-Houzel, S. (2009). The human brain in numbers: a linearly scaled-up primate brain. *Front Hum Neurosci* **3**: 31.
Hernandez-Miranda, L. R., A. Cariboni, C. Faux*, et al.* (2011). Robo1 regulates semaphorin signaling to guide the migration of cortical interneurons through the ventral forebrain. *J Neurosci* **31**(16): 6174-6187.
Hernandez, A. I., J. Wolk, J. Y. Hu*, et al.* (2009). Poly-(ADP-ribose) polymerase-1 is necessary for long-term facilitation in Aplysia. *J Neurosci* **29**(30): 9553-9562.
Hildebrandt, F., T. Benzing and N. Katsanis (2011). Ciliopathies. *N Engl J Med* **364**(16): 1533-1543.

Hindorff, L. A., Junkins, H.A., Hall, P.N., Mehta, J.P., and Manolio, T.A. A Catalog of Published Genome-Wide Association Studies. Retrieved 21.09.2011, from www.genome.gov/gwastudies.

Holopainen, L., T. Ahonen and H. Lyytinen (2001). Predicting delay in reading achievement in a highly transparent language. *J Learn Disabil* **34**(5): 401-413.

Howell, B. W., T. M. Herrick and J. A. Cooper (1999). Reelin-induced tyrosine [corrected] phosphorylation of disabled 1 during neuronal positioning. *Genes Dev* **13**(6): 643-648.

Hsiung, G. Y., B. J. Kaplan, T. L. Petryshen*, et al.* (2004). A dyslexia susceptibility locus (DYX7) linked to dopamine D4 receptor (DRD4) region on chromosome 11p15.5. *Am J Med Genet B Neuropsychiatr Genet* **125B**(1): 112-119.

Huang da, W., B. T. Sherman and R. A. Lempicki (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**(1): 44-57.

Huang da, W., B. T. Sherman, X. Zheng*, et al.* (2009). Extracting biological meaning from large gene lists with DAVID. *Curr Protoc Bioinformatics* **Chapter 13**: Unit 13 11.

Huang, Z. (2009). Molecular regulation of neuronal migration during neocortical development. *Mol Cell Neurosci* **42**(1): 11-22.

Humphreys, P., W. E. Kaufmann and A. M. Galaburda (1990). Developmental dyslexia in women: neuropathological findings in three patients. *Ann Neurol* **28**(6): 727-738.

Huntzinger, E. and E. Izaurralde (2011). Gene silencing by microRNAs: contributions of translational repression and mRNA decay. *Nat Rev Genet* **12**(2): 99-110.

Huth, J., M. Buchholz, J. M. Kraus*, et al.* (2010). Significantly improved precision of cell migration analysis in time-lapse video microscopy through use of a fully automated tracking system. *BMC Cell Biol* **11**: 24.

Hynd, G. W., J. Hall, E. S. Novey*, et al.* (1995). Dyslexia and corpus callosum morphology. *Arch Neurol* **52**(1): 32-38.

Hynd, G. W., M. Semrud-Clikeman, A. R. Lorys*, et al.* (1990). Brain morphology in developmental dyslexia and attention deficit disorder/hyperactivity. *Arch Neurol* **47**(8): 919-926.

Igo, R. P., Jr., N. H. Chapman, V. W. Berninger*, et al.* (2006). Genomewide scan for real-word reading subphenotypes of dyslexia: novel chromosome 13 locus and genetic complexity. *Am J Med Genet B Neuropsychiatr Genet* **141B**(1): 15-27.

Kaminen, N., K. Hannula-Jouppi, M. Kestila*, et al.* (2003). A genome scan for developmental dyslexia confirms linkage to chromosome 2p11 and suggests a new locus on 7q32. *J Med Genet* **40**(5): 340-345.

Kaplan, D. E., J. Gayan, J. Ahn*, et al.* (2002). Evidence for linkage and association with reading disability on 6p21.3-22. *Am J Hum Genet* **70**(5): 1287-1298.

Katusic, S. K., R. C. Colligan, W. J. Barbaresi*, et al.* (2001). Incidence of reading disability in a population-based birth cohort, 1976-1982, Rochester, Minn. *Mayo Clin Proc* **76**(11): 1081-1092.

Katzenellenbogen, B. S., M. M. Montano, T. R. Ediger*, et al.* (2000). Estrogen receptors: selective ligands, partners, and distinctive pharmacology. *Recent Prog Horm Res* **55**: 163-193; discussion 194-165.

Kauppinen, T. M. and R. A. Swanson (2007). The role of poly(ADP-ribose) polymerase-1 in CNS disease. *Neuroscience* **145**(4): 1267-1272.

Kawabe, H. and N. Brose (2010). The ubiquitin E3 ligase Nedd4-1 controls neurite development. *Cell Cycle* **9**(13): 2477-2478.

Kidd, T., K. Brose, K. J. Mitchell*, et al.* (1998). Roundabout controls axon crossing of the CNS midline and defines a novel subfamily of evolutionarily conserved guidance receptors. *Cell* **92**(2): 205-215.

Kim, Y. J., J. W. Huh, D. S. Kim*, et al.* (2009). Molecular characterization of the DYX1C1 gene and its application as a cancer biomarker. *J Cancer Res Clin Oncol* **135**(2): 265-270.

King, W. M., S. A. Giess and L. J. Lombardino (2007). Subtyping of children with developmental dyslexia via bootstrap aggregated clustering and the gap statistic: comparison with the double-deficit hypothesis. *Int J Lang Commun Disord* **42**(1): 77-95.

Kraus, W. L. and J. T. Lis (2003). PARP goes transcription. *Cell* **113**(6): 677-683.

Kriegstein, A. R. and S. C. Noctor (2004). Patterns of neuronal migration in the embryonic cortex. *Trends Neurosci* **27**(7): 392-399.

Ku, C. S., N. Naidoo and Y. Pawitan (2011). Revisiting Mendelian disorders through exome sequencing. *Hum Genet* **129**(4): 351-370.
Kumar, R. A., D. T. Pilz, T. D. Babatz*, et al.* (2010). TUBA1A mutations cause wide spectrum lissencephaly (smooth brain) and suggest that multiple neuronal migration pathways converge on alpha tubulins. *Hum Mol Genet* **19**(14): 2817-2827.
Kushch, A., K. Gross-Glenn, B. Jallad*, et al.* (1993). Temporal lobe surface area measurements on MRI in normal and dyslexic readers. *Neuropsychologia* **31**(8): 811-821.
Lander, E. S., L. M. Linton, B. Birren*, et al.* (2001). Initial sequencing and analysis of the human genome. *Nature* **409**(6822): 860-921.
Lander, E. S. and N. J. Schork (1994). Genetic dissection of complex traits. *Science* **265**(5181): 2037-2048.
Landerl, K., H. Wimmer and U. Frith (1997). The impact of orthographic consistency on dyslexia: a German-English comparison. *Cognition* **63**(3): 315-334.
Larsen, J. P., T. Hoien, I. Lundberg and H. Odegaard (1990). MRI evaluation of the size and symmetry of the planum temporale in adolescents with developmental dyslexia. *Brain Lang* **39**(2): 289-301.
Lawler, A. (2001). Archaeology. Writing gets a rewrite. *Science* **292**(5526): 2418-2420.
Leonard, C. M., K. K. Voeller, L. J. Lombardino*, et al.* (1993). Anomalous cerebral structure in dyslexia revealed with magnetic resonance imaging. *Arch Neurol* **50**(5): 461-469.
Levecque, C., A. Velayos-Baeza, Z. G. Holloway and A. P. Monaco (2009). The dyslexia-associated protein KIAA0319 interacts with adaptor protein 2 and follows the classical clathrin-mediated endocytosis pathway. *Am J Physiol Cell Physiol* **297**(1): C160-168.
Lewis, B. A., A. A. Avrich, L. A. Freebairn*, et al.* (2011). Literacy Outcomes of Children with Early Childhood Speech Sound Disorders: Impact of Endophenotypes. *J Speech Lang Hear Res*.
Lewitter, F. I., J. C. DeFries and R. C. Elston (1980). Genetic models of reading disability. *Behav Genet* **10**(1): 9-30.
Levy, S., G. Sutton, P. C. Ng*, et al.* (2007). The diploid genome sequence of an individual human. *PLoS Biol* **5**(10): e254.
Li, L. and J. R. Davie (2010). The role of Sp1 and Sp3 in normal and cancer cell biology. *Ann Anat* **192**(5): 275-283.
Lim, C. K., C. S. Ho, C. H. Chou and M. M. Waye (2011). Association of the rs3743205 variant of DYX1C1 with dyslexia in Chinese children. *Behav Brain Funct* **7**: 16.
Lin, P. I., J. M. Vance, M. A. Pericak-Vance and E. R. Martin (2007). No gene is an island: the flip-flop phenomenon. *Am J Hum Genet* **80**(3): 531-538.
Lind, P. A., M. Luciano, M. J. Wright*, et al.* (2010). Dyslexia and DCDC2: normal variation in reading and spelling is associated with DCDC2 polymorphisms in an Australian population sample. *Eur J Hum Genet* **18**(6): 668-673.
Litt, J., H. G. Taylor, N. Klein and M. Hack (2005). Learning disabilities in children with very low birthweight: prevalence, neuropsychological correlates, and educational interventions. *J Learn Disabil* **38**(2): 130-141.
Liu, J. S. (2011). Molecular genetics of neuronal migration disorders. *Curr Neurol Neurosci Rep* **11**(2): 171-178.
Liu, L., N. Xie, P. Rennie*, et al.* (2011). Consensus PP1 binding motifs regulate transcriptional corepression and alternative RNA splicing activities of the steroid receptor coregulators, p54nrb and PSF. *Mol Endocrinol* **25**(7): 1197-1210.
Livak, K. J. and T. D. Schmittgen (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**(4): 402-408.
Lock, J. G., B. Wehrle-Haller and S. Stromblad (2008). Cell-matrix adhesion complexes: master control machinery of cell migration. *Semin Cancer Biol* **18**(1): 65-76.
LoTurco, J. J. and J. Bai (2006). The multipolar stage and disruptions in neuronal migration. *Trends Neurosci* **29**(7): 407-413.
Lowery, L. A., J. Rubin and H. Sive (2007). Whitesnake/sfpq is required for cell survival and neuronal development in the zebrafish. *Dev Dyn* **236**(5): 1347-1357.
Luciano, M., P. A. Lind, D. L. Duffy*, et al.* (2007). A haplotype spanning KIAA0319 and TTRAP is associated with normal variation in reading and spelling ability. *Biol Psychiatry* **62**(7): 811-817.
Lyytinen, H., M. Aro, K. Eklund*, et al.* (2004). The development of children at familial risk for dyslexia: birth to early school age. *Ann Dyslexia* **54**(2): 184-220.

Ma, K., S. Zheng and Z. Zuo (2006). The transcription factor regulatory factor X1 increases the expression of neuronal glutamate transporter type 3. *J Biol Chem* **281**(30): 21250-21255.
MacQuarrie, K. L., A. P. Fong, R. H. Morse and S. J. Tapscott (2011). Genome-wide transcription factor binding: beyond direct target regulation. *Trends Genet* **27**(4): 141-148.
Malaterre, J., T. Mantamadiotis, S. Dworkin*, et al.* (2008). c-Myb is required for neural progenitor cell proliferation and maintenance of the neural stem cell niche in adult brain. *Stem Cells* **26**(1): 173-181.
Manzini, M. C. and C. A. Walsh (2011). What disorders of cortical development tell us about the cortex: one plus one does not always make two. *Curr Opin Genet Dev* **21**(3): 333-339.
Marin, O. and J. L. Rubenstein (2001). A long, remarkable journey: tangential migration in the telencephalon. *Nat Rev Neurosci* **2**(11): 780-790.
Marino, C., A. Citterio, R. Giorda*, et al.* (2007). Association of short-term memory with a variant within DYX1C1 in developmental dyslexia. *Genes Brain Behav* **6**(7): 640-646.
Marino, C., R. Giorda, M. Luisa Lorusso*, et al.* (2005). A family-based association study does not support DYX1C1 on 15q21.3 as a candidate gene in developmental dyslexia. *Eur J Hum Genet* **13**(4): 491-499.
Marino, C., R. Giorda, L. Vanzin*, et al.* (2004). A locus on 15q15-15qter influences dyslexia: further support from a transmission/disequilibrium study in an Italian speaking population. *J Med Genet* **41**(1): 42-46.
Marino, C., S. Mascheretti, V. Riva*, et al.* (2011). Pleiotropic effects of DCDC2 and DYX1C1 genes on language and mathematics traits in nuclear families of developmental dyslexia. *Behav Genet* **41**(1): 67-76.
Marino, C., H. Meng, S. Mascheretti*, et al.* (2011). DCDC2 genetic variants and susceptibility to developmental dyslexia. *Psychiatr Genet*.
Marlow, A. J., S. E. Fisher, C. Francks*, et al.* (2003). Use of multivariate linkage analysis for dissection of a complex cognitive trait. *Am J Hum Genet* **72**(3): 561-570.
Marth, G. T., F. Yu, A. R. Indap*, et al.* (2011). The functional spectrum of low-frequency coding variation. *Genome Biol* **12**(9): R84.
Martinez-Cerdeno, V., S. C. Noctor and A. R. Kriegstein (2006). Estradiol stimulates progenitor cell division in the ventricular and subventricular zones of the embryonic neocortex. *Eur J Neurosci* **24**(12): 3475-3488.
Massinen, S., M.-E. Hokkanen, H. Matsson*, et al.* (2011). Increased Expression of the Dyslexia Candidate Gene DCDC2 Affects Length and Signaling of Primary Cilia in Neurons. *PLoS One* **6**(6): e20580.
Matsson, H., K. Tammimies, M. Zucchelli*, et al.* (2011). SNP variations in the 7q33 region containing DGKI are associated with dyslexia in the Finnish and German populations. *Behav Genet* **41**(1): 134-140.
Matsuo, K., T. Nishimura, A. Hayakawa*, et al.* (2010). Involvement of a centrosomal protein kendrin in the maintenance of centrosome cohesion by modulating Nek2A kinase activity. *Biochem Biophys Res Commun* **398**(2): 217-223.
Mattick, J. S. (2011). The central role of RNA in human development and cognition. *FEBS Lett* **585**(11): 1600-1616.
McCarthy, M. I., G. R. Abecasis, L. R. Cardon*, et al.* (2008). Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* **9**(5): 356-369.
McCarthy, M. M. (2008). Estradiol and the developing brain. *Physiol Rev* **88**(1): 91-124.
Meda, S. A., J. Gelernter, J. R. Gruen*, et al.* (2008). Polymorphism of DCDC2 Reveals Differences in Cortical Morphology of Healthy Individuals-A Preliminary Voxel Based Morphometry Study. *Brain Imaging Behav* **2**(1): 21-26.
Meng, H., K. Hager, M. Held*, et al.* (2005). TDT-association analysis of EKN1 and dyslexia in a Colorado twin cohort. *Hum Genet* **118**(1): 87-90.
Meng, H., N. R. Powers, L. Tang*, et al.* (2011). A dyslexia-associated variant in DCDC2 changes gene expression. *Behav Genet* **41**(1): 58-66.
Meng, H., S. D. Smith, K. Hager*, et al.* (2005). DCDC2 is associated with reading disability and modulates neuronal development in the brain. *Proc Natl Acad Sci U S A* **102**(47): 17053-17058.

Meyer, R., S. S. Wolf and M. Obendorf (2007). PRMT2, a member of the protein arginine methyltransferase family, is a coactivator of the androgen receptor. *J Steroid Biochem Mol Biol* **107**(1-2): 1-14.
Mills, R. E., K. Walter, C. Stewart*, et al.* (2011). Mapping copy number variation by population-scale genome sequencing. *Nature* **470**(7332): 59-65.
Morris, D. W., L. Robinson, D. Turic*, et al.* (2000). Family-based association mapping provides evidence for a gene for reading disability on chromosome 15q. *Hum Mol Genet* **9**(5): 843-848.
Moskvina, V., N. Craddock, P. Holmans*, et al.* (2009). Gene-wide analyses of genome-wide association data sets: evidence for multiple common risk alleles for schizophrenia and bipolar disorder and for overlap in genetic risk. *Mol Psychiatry* **14**(3): 252-260.
Naftolin, F., N. J. MacLusky, C. Z. Leranth*, et al.* (1988). The cellular effects of estrogens on neuroendocrine tissues. *J Steroid Biochem* **30**(1-6): 195-207.
Nagano, T., S. Morikubo and M. Sato (2004). Filamin A and FILIP (Filamin A-Interacting Protein) regulate cell polarity and motility in neocortical subventricular and intermediate zones during radial migration. *J Neurosci* **24**(43): 9648-9657.
Nakanishi, K., Y. Tokita, S. Aono*, et al.* (2010). Neuroglycan C, a brain-specific chondroitin sulfate proteoglycan, interacts with pleiotrophin, a heparin-binding growth factor. *Neurochem Res* **35**(8): 1131-1137.
Narlikar, L. and I. Ovcharenko (2009). Identifying regulatory elements in eukaryotic genomes. *Brief Funct Genomic Proteomic* **8**(4): 215-230.
Newbury, D. F., S. Paracchini, T. S. Scerri*, et al.* (2011). Investigation of dyslexia and SLI risk variants in reading- and language-impaired subjects. *Behav Genet* **41**(1): 90-104.
Nicolson, R. I. a. F., A.J. (2008).Dyslexia, Learning and the Brain. The MIT Press.
Nopola-Hemmi, J., B. Myllyluoma, T. Haltia*, et al.* (2001). A dominant gene for developmental dyslexia on chromosome 3. *J Med Genet* **38**(10): 658-664.
Nopola-Hemmi, J., M. Taipale, T. Haltia*, et al.* (2000). Two translocations of chromosome 15q associated with dyslexia. *J Med Genet* **37**(10): 771-775.
Nothen, M. M., G. Schulte-Korne, T. Grimm*, et al.* (1999). Genetic linkage analysis with dyslexia: evidence for linkage of spelling disability to chromosome 15. *Eur Child Adolesc Psychiatry* **8 Suppl 3**: 56-59.
O'Brien, T. W., B. J. O'Brien and R. A. Norman (2005). Nuclear MRP genes and mitochondrial disease. *Gene* **354**: 147-151.
Ogura, Y., M. Azuma, Y. Tsuboi*, et al.* (2006). TFII-I down-regulates a subset of estrogen-responsive genes through its interaction with an initiator element and estrogen receptor alpha. *Genes Cells* **11**(4): 373-381.
Olson, R. K. (2002). Dyslexia: nature and nurture. *Dyslexia* **8**(3): 143-159.
Osborne, C. K., K. Hobbs and J. M. Trent (1987). Biological differences among MCF-7 human breast cancer cell lines from different laboratories. *Breast Cancer Res Treat* **9**(2): 111-121.
Osterlund, M. K. and Y. L. Hurd (2001). Estrogen receptors in the human forebrain and the relation to neuropsychiatric disorders. *Prog Neurobiol* **64**(3): 251-267.
Pahlman, S., A. I. Ruusala, L. Abrahamsson*, et al.* (1984). Retinoic acid-induced differentiation of cultured human neuroblastoma cells: a comparison with phorbolester-induced differentiation. *Cell Differ* **14**(2): 135-144.
Pakkenberg, B. and H. J. Gundersen (1997). Neocortical neuron number in humans: effect of sex and age. *J Comp Neurol* **384**(2): 312-320.
Paracchini, S. (2011). Dissection of genetic associations with language-related traits in population-based cohorts. *J Neurodev Disord*.
Paracchini, S., Q. W. Ang, F. J. Stanley*, et al.* (2011). Analysis of dyslexia candidate genes in the Raine cohort representing the general Australian population. *Genes Brain Behav* **10**(2): 158-165.
Paracchini, S., C. D. Steer, L. L. Buckingham*, et al.* (2008). Association of the KIAA0319 dyslexia susceptibility gene with reading skills in the general population. *Am J Psychiatry* **165**(12): 1576-1584.
Paracchini, S., A. Thomas, S. Castro*, et al.* (2006). The chromosome 6p22 haplotype associated with dyslexia reduces the expression of KIAA0319, a novel gene involved in neuronal migration. *Hum Mol Genet* **15**(10): 1659-1666.
Patton, J. G., E. B. Porro, J. Galceran*, et al.* (1993). Cloning and characterization of PSF, a novel pre-mRNA splicing factor. *Genes Dev* **7**(3): 393-406.

Paulesu, E., J. F. Demonet, F. Fazio, *et al.* (2001). Dyslexia: cultural diversity and biological unity. *Science* **291**(5511): 2165-2167.
Pennington, B. F. (1990). The genetics of dyslexia. *J Child Psychol Psychiatry* **31**(2): 193-201.
Pennington, B. F. (1995). Genetics of learning disabilities. *J Child Neurol* **10 Suppl 1**: S69-77.
Pennington, B. F., J. W. Gilger, D. Pauls, *et al.* (1991). Evidence for major gene transmission of developmental dyslexia. *JAMA* **266**(11): 1527-1534.
Pennington, B. F., S. D. Smith, W. J. Kimberling, *et al.* (1987). Left-handedness and immune disorders in familial dyslexics. *Arch Neurol* **44**(6): 634-639.
Perez Jurado, L. A., Y. K. Wang, R. Peoples, *et al.* (1998). A duplicated gene in the breakpoint regions of the 7q11.23 Williams-Beuren syndrome deletion encodes the initiator binding protein TFII-I and BAP-135, a phosphorylation target of BTK. *Hum Mol Genet* **7**(3): 325-334.
Perou, C. M. and A. L. Borresen-Dale (2011). Systems biology and genomics of breast cancer. *Cold Spring Harb Perspect Biol* **3**(2).
Peschansky, V. J., T. J. Burbridge, A. J. Volz, *et al.* (2009). The Effect of Variation in Expression of the Candidate Dyslexia Susceptibility Gene Homolog Kiaa0319 on Neuronal Migration and Dendritic Morphology in the Rat. *Cereb Cortex*.
Petryshen, T. L., B. J. Kaplan, M. Fu Liu, *et al.* (2001). Evidence for a susceptibility locus on chromosome 6q influencing phonological coding dyslexia. *Am J Med Genet* **105**(6): 507-517.
Petryshen, T. L., B. J. Kaplan, M. L. Hughes, *et al.* (2002). Supportive evidence for the DYX3 dyslexia susceptibility gene in Canadian families. *J Med Genet* **39**(2): 125-126.
Peyrard-Janvid, M., H. Anthoni, P. Onkamo, *et al.* (2004). Fine mapping of the 2p11 dyslexia locus and exclusion of TACR1 as a candidate gene. *Hum Genet* **114**(5): 510-516.
Platko, J. V., F. B. Wood, I. Pelser, *et al.* (2008). Association of reading disability on chromosome 6p22 in the Afrikaner population. *Am J Med Genet B Neuropsychiatr Genet* **147B**(7): 1278-1287.
Poelmans, G., J. J. Engelen, J. Van Lent-Albrechts, *et al.* (2009). Identification of novel dyslexia candidate genes through the analysis of a chromosomal deletion. *Am J Med Genet B Neuropsychiatr Genet* **150B**(1): 140-147.
Pramparo, T., O. Libiger, S. Jain, *et al.* (2011). Global developmental gene expression and pathway analysis of normal brain development and mouse models of human neuronal migration defects. *PLoS Genet* **7**(3): e1001331.
Price, C. J. (2000). The anatomy of language: contributions from functional neuroimaging. *J Anat* **197 Pt 3**: 335-359.
Puolakanaho, A., T. Ahonen, M. Aro, *et al.* (2008). Developmental links of very early phonological and language skills to second grade reading outcomes: strong to accuracy but only minor to fluency. *J Learn Disabil* **41**(4): 353-370.
Qi, C., J. Chang, Y. Zhu, *et al.* (2002). Identification of protein arginine methyltransferase 2 as a coactivator for estrogen receptor alpha. *J Biol Chem* **277**(32): 28624-28630.
Rabin, M., X. L. Wen, M. Hepburn, *et al.* (1993). Suggestive linkage of developmental dyslexia to chromosome 1p34-p36. *Lancet* **342**(8864): 178.
Ramsay, R. G. and T. J. Gonda (2008). MYB function in normal and cancer cells. *Nat Rev Cancer* **8**(7): 523-534.
Ramus, F. and G. Szenkovits (2008). What phonological deficit? *Q J Exp Psychol (Colchester)* **61**(1): 129-141.
Raskind, W. H., R. P. Igo, N. H. Chapman, *et al.* (2005). A genome scan in multigenerational families with dyslexia: Identification of a novel locus on chromosome 2q that contributes to phonological decoding efficiency. *Mol Psychiatry* **10**(7): 699-711.
Ravasi, T., H. Suzuki, C. V. Cannistraci, *et al.* (2010). An atlas of combinatorial transcriptional regulation in mouse and man. *Cell* **140**(5): 744-752.
Raz, L., M. M. Khan, V. B. Mahesh, *et al.* (2008). Rapid estrogen signaling in the brain. *Neurosignals* **16**(2-3): 140-153.
Reich, D. E. and E. S. Lander (2001). On the allelic spectrum of human disease. *Trends Genet* **17**(9): 502-510.
Richardson, S. O. (1992). Historical perspectives on dyslexia. *J Learn Disabil* **25**(1): 40-47; discussion 48-52.
Risch, N. and K. Merikangas (1996). The future of genetic studies of complex human diseases. *Science* **273**(5281): 1516-1517.

Risch, N. J. (2000). Searching for genetic determinants in the new millennium. *Nature* **405**(6788): 847-856.
Rissman, E. F., A. L. Heck, J. E. Leonard*, et al.* (2002). Disruption of estrogen receptor beta gene impairs spatial learning in female mice. *Proc Natl Acad Sci U S A* **99**(6): 3996-4001.
Roeske, D., K. U. Ludwig, N. Neuhoff*, et al.* (2011). First genome-wide association scan on neurophysiological endophenotypes points to trans-regulation effects on SLC2A3 in dyslexic children. *Mol Psychiatry* **16**(1): 97-107.
Rosen, G. D., J. Bai, Y. Wang*, et al.* (2007). Disruption of neuronal migration by RNAi of Dyx1c1 results in neocortical and hippocampal malformations. *Cereb Cortex* **17**(11): 2562-2572.
Roy, A. L., M. Meisterernst, P. Pognonec and R. G. Roeder (1991). Cooperative interaction of an initiator-binding transcription initiation factor and the helix-loop-helix activator USF. *Nature* **354**(6350): 245-248.
Rumsey, J. M., M. Casanova, G. B. Mannheim*, et al.* (1996). Corpus callosum morphology, as measured with MRI, in dyslexic men. *Biol Psychiatry* **39**(9): 769-775.
Rutter, M., A. Caspi, D. Fergusson*, et al.* (2004). Sex differences in developmental reading disability: new findings from 4 epidemiological studies. *Jama* **291**(16): 2007-2012.
Rutter, M. and W. Yule (1975). The concept of specific reading retardation. *J Child Psychol Psychiatry* **16**(3): 181-197.
Sakabe, N. J. and M. A. Nobrega (2010). Genome-wide maps of transcription regulatory elements. *Wiley Interdiscip Rev Syst Biol Med* **2**(4): 422-437.
Salmelin, R., E. Service, P. Kiesila*, et al.* (1996). Impaired visual word processing in dyslexia revealed with magnetoencephalography. *Ann Neurol* **40**(2): 157-162.
Saviour, P., S. Kumar, U. Kiran*, et al.* (2008). Allelic variants of DYX1C1 are not associated with dyslexia in India. *Indian J Hum Genet* **14**(3): 99-102.
Scerri, T. S., S. E. Fisher, C. Francks*, et al.* (2004). Putative functional alleles of DYX1C1 are not associated with dyslexia susceptibility in a large sample of sibling pairs from the UK. *J Med Genet* **41**(11): 853-857.
Scerri, T. S., A. P. Morris, L. L. Buckingham*, et al.* (2011). DCDC2, KIAA0319 and CMIP are associated with reading-related traits. *Biol Psychiatry* **70**(3): 237-245.
Scerri, T. S., S. Paracchini, A. Morris*, et al.* (2010). Identification of candidate genes for dyslexia susceptibility on chromosome 18. *PLoS One* **5**(10): e13712.
Schaar, B. T., K. Kinoshita and S. K. McConnell (2004). Doublecortin microtubule affinity is regulated by a balance of kinase and phosphatase activity at the leading edge of migrating neurons. *Neuron* **41**(2): 203-213.
Schulte-Korne, G., T. Grimm, M. M. Nothen*, et al.* (1998). Evidence for linkage of spelling disability to chromosome 15. *Am J Hum Genet* **63**(1): 279-282.
Schumacher, J., H. Anthoni, F. Dahdouh*, et al.* (2006). Strong genetic evidence of DCDC2 as a susceptibility gene for dyslexia. *Am J Hum Genet* **78**(1): 52-62.
Schumacher, J., I. R. Konig, T. Schroder*, et al.* (2008). Further evidence for a susceptibility locus contributing to reading disability on chromosome 15q15-q21. *Psychiatr Genet* **18**(3): 137-142.
Seeger, T., I. Fedorova, F. Zheng*, et al.* (2004). M2 muscarinic acetylcholine receptor knock-out mice show deficits in behavioral flexibility, working memory, and hippocampal plasticity. *J Neurosci* **24**(45): 10117-10127.
Serrano, F. and S. Defior (2008). Dyslexia speed problems in a transparent orthography. *Ann Dyslexia* **58**(1): 81-95.
Shaywitz, B. A., S. E. Shaywitz, K. R. Pugh*, et al.* (2002). Disruption of posterior brain systems for reading in children with developmental dyslexia. *Biol Psychiatry* **52**(2): 101-110.
Shaywitz, S. E. (1998). Dyslexia. *N Engl J Med* **338**(5): 307-312.
Shaywitz, S. E., M. D. Escobar, B. A. Shaywitz*, et al.* (1992). Evidence that dyslexia may represent the lower tail of a normal distribution of reading ability. *N Engl J Med* **326**(3): 145-150.
Shaywitz, S. E. and B. A. Shaywitz (2008). Paying attention to reading: the neurobiology of reading and dyslexia. *Dev Psychopathol* **20**(4): 1329-1349.
Shaywitz, S. E., B. A. Shaywitz, K. R. Pugh*, et al.* (1998). Functional disruption in the organization of the brain for reading in dyslexia. *Proc Natl Acad Sci U S A* **95**(5): 2636-2641.

Silani, G., U. Frith, J. F. Demonet, *et al.* (2005). Brain abnormalities underlying altered activation in dyslexia: a voxel based morphometry study. *Brain* **128**(Pt 10): 2453-2461.
Skiba, T., N. Landi, R. Wagner and E. L. Grigorenko (2011). In search of the perfect phenotype: an analysis of linkage and association studies of reading and reading-related processes. *Behav Genet* **41**(1): 6-30.
Smith, S. D., W. J. Kimberling and B. F. Pennington (1991). Screening for multiple genes influencing dyslexia. *Reading and Writing* **3**(3): 285-298.
Smith, S. D., W. J. Kimberling, B. F. Pennington and H. A. Lubs (1983). Specific reading disability: identification of an inherited form through linkage analysis. *Science* **219**(4590): 1345-1347.
Smyth, G. K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* **3**: Article3.
Soderberg, O., M. Gullberg, M. Jarvius, *et al.* (2006). Direct observation of individual endogenous protein complexes in situ by proximity ligation. *Nat Methods* **3**(12): 995-1000.
Steensgaard, P., M. Garre, I. Muradore, *et al.* (2004). Sgt1 is required for human kinetochore assembly. *EMBO Rep* **5**(6): 626-631.
Stein, C. M., J. H. Schick, H. Gerry Taylor, *et al.* (2004). Pleiotropic effects of a chromosome 3 locus on speech-sound disorder and reading. *Am J Hum Genet* **74**(2): 283-297.
Stein, J. (2001). The magnocellular theory of developmental dyslexia. *Dyslexia* **7**(1): 12-36.
Steiner, J., B. Bogerts, M. L. Schroeter and H. G. Bernstein (2011). S100B protein in neurodegenerative disorders. *Clin Chem Lab Med* **49**(3): 409-424.
Stoodley, C. J. and J. F. Stein (2011). The cerebellum and dyslexia. *Cortex* **47**(1): 101-116.
Swoboda, P., H. T. Adler and J. H. Thomas (2000). The RFX-type transcription factor DAF-19 regulates sensory neuron cilium formation in C. elegans. *Mol Cell* **5**(3): 411-421.
Szalkowski, C. E., J. R. Hinman, S. W. Threlkeld, *et al.* (2011). Persistent spatial working memory deficits in rats following in utero RNAi of Dyx1c1. *Genes Brain Behav* **10**(2): 244-252.
Taipale, M., N. Kaminen, J. Nopola-Hemmi, *et al.* (2003). A candidate gene for developmental dyslexia encodes a nuclear tetratricopeptide repeat domain protein dynamically regulated in brain. *Proc Natl Acad Sci U S A* **100**(20): 11553-11558.
Tateishi, Y., Y. Kawabe, T. Chiba, *et al.* (2004). Ligand-dependent switching of ubiquitin-proteasome pathways for estrogen receptor. *Embo J* **23**(24): 4813-4823.
Tateishi, Y., R. Sonoo, Y. Sekiya, *et al.* (2006). Turning off estrogen receptor beta-mediated transcription requires estrogen-dependent receptor proteolysis. *Mol Cell Biol* **26**(21): 7966-7976.
Temple, E., R. A. Poldrack, J. Salidis, *et al.* (2001). Disrupted neural responses to phonological and orthographic processing in dyslexic children: an fMRI study. *Neuroreport* **12**(2): 299-307.
Thompson, J. D., T. J. Gibson, F. Plewniak, *et al.* (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**(24): 4876-4882.
Threlkeld, S. W., M. M. McClure, J. Bai, *et al.* (2007). Developmental disruptions and behavioral impairments in rats following in utero RNAi of Dyx1c1. *Brain Res Bull* **71**(5): 508-514.
Tsai, J. W., K. H. Bremner and R. B. Vallee (2007). Dual subcellular roles for LIS1 and dynein in radial neuronal migration in live brain tissue. *Nat Neurosci* **10**(8): 970-979.
Tsai, L. H. and J. G. Gleeson (2005). Nucleokinesis in neuronal migration. *Neuron* **46**(3): 383-388.
Turic, D., L. Robinson, M. Duke, *et al.* (2003). Linkage disequilibrium mapping provides further evidence of a gene for reading disability on chromosome 6p21.3-22. *Mol Psychiatry* **8**(2): 176-185.
Turkeltaub, P. E., L. Gareau, D. L. Flowers, *et al.* (2003). Development of neural mechanisms for reading. *Nat Neurosci* **6**(7): 767-773.
Tzenova, J., B. J. Kaplan, T. L. Petryshen and L. L. Field (2004). Confirmation of a dyslexia susceptibility locus on chromosome 1p34-p36 in a set of 100 Canadian families. *Am J Med Genet B Neuropsychiatr Genet* **127B**(1): 117-124.
Uribe, V. (2010). The beta-tubulin gene TUBB2B is involved in a large spectrum of neuronal migration disorders. *Clin Genet* **77**(1): 34-35.

Wang, K., M. Li and H. Hakonarson (2010). Analysing biological pathways in genome-wide association studies. *Nat Rev Genet* **11**(12): 843-854.
Wang, L., S. Andersson, M. Warner and J. A. Gustafsson (2001). Morphological abnormalities in the brains of estrogen receptor beta knockout mice. *Proc Natl Acad Sci U S A* **98**(5): 2792-2796.
Wang, L., S. Andersson, M. Warner and J. A. Gustafsson (2003). Estrogen receptor (ER)beta knockout mice reveal a role for ERbeta in migration of cortical neurons in the developing brain. *Proc Natl Acad Sci U S A* **100**(2): 703-708.
Wang, Y., M. Paramasivam, A. Thomas*, et al.* (2006). DYX1C1 functions in neuronal migration in developing neocortex. *Neuroscience* **143**(2): 515-522.
Wang, Y., X. Yin, G. Rosen*, et al.* (2011). Dcdc2 knockout mice display exacerbated developmental disruptions following knockdown of doublecortin. *Neuroscience* **190**: 398-408.
Wapinski, O. and H. Y. Chang (2011). Long noncoding RNAs and human disease. *Trends in Cell Biology* **21**(6): 354-361.
Vasudevan, N. and D. W. Pfaff (2008). Non-genomic actions of estrogens and their interaction with genomic actions in the brain. *Front Neuroendocrinol* **29**(2): 238-257.
Watson, J. D. and F. H. Crick (1953). The structure of DNA. *Cold Spring Harb Symp Quant Biol* **18**: 123-131.
Velayos-Baeza, A., C. Levecque, K. Kobayashi*, et al.* (2010). The dyslexia-associated KIAA0319 protein undergoes proteolytic processing with {gamma}-secretase-independent intramembrane cleavage. *J Biol Chem* **285**(51): 40148-40162.
Velayos-Baeza, A., C. Toma, S. da Roza*, et al.* (2007). Alternative splicing in the dyslexia-associated gene KIAA0319. *Mamm Genome* **18**(9): 627-634.
Velayos-Baeza, A., C. Toma, S. Paracchini and A. P. Monaco (2008). The dyslexia-associated gene KIAA0319 encodes highly N- and O-glycosylated plasma membrane and secreted isoforms. *Hum Mol Genet* **17**(6): 859-871.
Venkatesh, S. K., A. Siddaiah, P. Padakannaya and N. B. Ramachandra (2011). An examination of candidate gene SNPs for dyslexia in an Indian sample. *Behav Genet* **41**(1): 105-109.
Venter, J. C., M. D. Adams, E. W. Myers*, et al.* (2001). The sequence of the human genome. *Science* **291**(5507): 1304-1351.
Vernes, S. C., P. L. Oliver, E. Spiteri*, et al.* (2011). Foxp2 regulates gene networks implicated in neurite outgrowth in the developing brain. *PLoS Genet* **7**(7): e1002145.
Wheeler, D. A., M. Srinivasan, M. Egholm*, et al.* (2008). The complete genome of an individual by massively parallel DNA sequencing. *Nature* **452**(7189): 872-876.
Vidal, M., M. E. Cusick and A. L. Barabasi (2011). Interactome networks and human disease. *Cell* **144**(6): 986-998.
Wigg, K. G., J. M. Couto, Y. Feng*, et al.* (2004). Support for EKN1 as the susceptibility locus for dyslexia on 15q21. *Mol Psychiatry* **9**(12): 1111-1121.
Viholainen, H., T. Ahonen, M. Cantell*, et al.* (2002). Development of early motor skills and language in children at risk for familial dyslexia. *Dev Med Child Neurol* **44**(11): 761-769.
Wilcke, A., C. Ligges, J. Burkhardt*, et al.* (2011). Imaging genetics of FOXP2 in dyslexia. *Eur J Hum Genet*.
Wilcke, A., J. Weissfuss, H. Kirsten*, et al.* (2009). The role of gene DCDC2 in German dyslexics. *Ann Dyslexia* **59**(1): 1-11.
Willcutt, E. G., B. F. Pennington and J. C. DeFries (2000). Twin study of the etiology of comorbidity between reading disability and attention-deficit/hyperactivity disorder. *Am J Med Genet* **96**(3): 293-301.
Willcutt, E. G., B. F. Pennington, L. Duncan*, et al.* (2010). Understanding the complex etiologies of developmental disorders: behavioral and molecular genetic approaches. *J Dev Behav Pediatr* **31**(7): 533-544.
Vogler, G. P., J. C. DeFries and S. N. Decker (1985). Family history as an indicator of risk for reading disability. *J Learn Disabil* **18**(7): 419-421.
von Plessen, K., A. Lundervold, N. Duta*, et al.* (2002). Less developed corpus callosum in dyslexic subjects--a structural MRI study. *Neuropsychologia* **40**(7): 1035-1044.
Vukovic, R. K. and L. S. Siegel (2006). The double-deficit hypothesis: a comprehensive analysis of the evidence. *J Learn Disabil* **39**(1): 25-47.

Yu, G., T. Zerucha, M. Ekker and J. L. Rubenstein (2001). Evidence that GRIP, a PDZ-domain protein which is expressed in the embryonic forebrain, co-activates transcription with DLX homeodomain proteins. *Brain Res Dev Brain Res* **130**(2): 217-230.

Zhang, B., S. Kirov and J. Snoddy (2005). WebGestalt: an integrated system for exploring gene sets in various biological contexts. *Nucleic Acids Res* **33**(Web Server issue): W741-748.

Zheng, M., T. Cierpicki, A. J. Burdette*, et al.* (2011). Structural Features and Chaperone Activity of the NudC Protein Family. *J Mol Biol* **409**(5): 722-741.

Ziegler, J. C., D. Bertrand, D. Toth*, et al.* (2010). Orthographic depth and its impact on universal predictors of reading: a cross-language investigation. *Psychol Sci* **21**(4): 551-559.