PhD degree in Molecular Medicine

(curriculum in Molecular Oncology and Human Genetics)

European School of Molecular Medicine (SEMM),

University of Milan and University of Naples "Federico II"

Settore disciplinare: Bio/18

A cell reprogramming-based approach to study 7q11.23 gene dosage imbalances in Williams Beuren syndrome and autism spectrum disorder

Sina Atashpaz (PharmD)

European Institute of Oncology (IEO), Milan

Matricola n. R09397

Supervisor: Prof. Giuseppe Testa, MD, PhD

European Institute of Oncology (IEO) and University of Milan, Italy

Added Supervisor: Prof. Pier Giuseppe Pelicci, MD, PhD

European Institute of Oncology (IEO) and University of Milan, Italy

Prof. Rick Livesey, MB Bchir, PhD

The Gurdon Institute and University of Cambridge, UK

Anno accademico 2013-2014

Dedication

To my devoted mother and supportive father

who has smoothed the path for me

to step in

walk

and even run

to achieve my goals

To all my true teachers, my devoted professors,

who not only taught me science

but

enlightened me of life

its ups and downs . . .

To whom I owe the leaping delight, my patient wife,

who thinks the same thoughts without need of speech

Acknowledgements

My special thanks go to my encouraging friends Thomas Burgold and Elena Signaroldi who have been always available to assist me during all steps of my project.

I would also like to thank my supervisors, Prof. G. Testa (European Institute of Oncology and University of Milan, Italy), Prof. P.G. Pelicci (European Institute of Oncology and University of Milan, Italy), and Prof. R. Livesey (The Gurdon Institute and University of Cambridge, UK) for their kind guidance and supervision during this project and my examiners Prof. Jens Christian Schwamborn (University of Luxembourg, Luxembourg) and Prof. Saverio Minucci (European Institute of Oncology and University of Milan, Italy) for their time, useful suggestions and excellent discussions during my PhD exam.

I would like to offer my special thanks to my dear friends: P. Germain, M. Zanella, A. Adamo, G. D'Agostino, V. Albertin, P. Lo Riso, G. Barbagiovanni, G. Fragola, J. Sgualdino, S. Buontempo, P. Tripathi, S. Cristofanon, M. Gabriele, E. Cesarini, E. de Nittis, A. Chronowska, A. Piunti, M. Doni, M. Dal Molin, L. Marelli, P. Laise, L. Chiapperino, M. Damjanovicova, V. Das, P. Fuentes and A. Vitriolo who have kindly helped me during this project.

I wish to thank to my Iranian friends at IFOM-IEO campus, Leila Dardaei, Betsabeh Khoramian Tusi, Mahshid Rahmat, Parinaz Mehdipour, Sara Rohban and Seyed amir Hosseini who helped me feel less far from home by their kind presence.

I would like to thank to all our collaborators: Josh Chenoweth and Ronald McKay (Lieber Institute for Brain Development, Baltimore, Maryland, USA), Lucia Micale, Carmela Fusco, Bartolomeo Augello, Orazio Palumbo, Massimo Carella, and Giuseppe Merla (Medical

Genetics Unit, IRCCS Casa Sollievo della Sofferenza Hospital, San Giovanni Rotondo, Italy), Christian Unger and Peter Andrews (Department of Biomedical Sciences, University of Sheffield, Sheffield, UK), Brad Hamilton (Stemgent, Cambridge, Massachusetts, USA), Emilio Donti, and Paolo Prontera (Medical Genetics Unit, Hospital Santa Maria della Misericordia, University of Perugia, Perugia, Italy), Giancarlo Pruneri1(Department of Experimental Oncology, European Institute of Oncology, Milan, Italy), Angelo Selicorni (Unità Operativa Semplice (UOS) Genetica Clinica Pediatrica, Fondazione Monza e Brianza per il Bambino e la sua Mamma (Fondazione MBBM), Azienda Ospedaliera San Gerardo, Monza, Italy), Gustavo Mostoslavsky and Andreia Gianotti Sommer (Center for Regenerative Medicine of Boston University, Boston, US), Federica Pisati (IFOM-IEO campus, Milan, Italy), Federica Ungaro and Massimiliano Caiazzo (San Raffaele Scientific Institute, Milan, Italy), Stefano Camnasio, Alessia Delli Carri and Elena Cattaneo (University of Milan, Milan. Italy).

I would like to thank to AFSW (Associazione Famiglie Sindrome di Williams) and AISW (Associazione Italiana Sindrome di Williams) for agreeing to participate and making this study possible and the Genomic and Genetic Disorder Biobank, Galliera Genetic Bank and members of the Telethon Network of Genetic Biobanks (project numbers GTB12001G and GTB12001A), along with the EuroBioBank network, for providing us with specimens, the Umberto Veronesi Foundation and University of Milan for the fellowship awarded to me.

Very special thanks go to kind staffs of SEMM office, Veronica Viscardi and Francesa Fiore who have kindly helped me during the days of need.

There is no word that could express my gratitude to Stefano Casola and Mio Sumie who have been always available to make a time for me during the days of need.

And at last but not least, I would like to thanks all the kind staffs of IFOM-IEO campus for their kind guidance and technical supports.

Table of content

List of Abbreviations

5-HT	Serotonin
7dup-ASD	7q11.23 microduplication associated to autistic spectrum disorder
aCGH	Array-based Comparative Genomic Hybridization
ADHD	Attention deficit hyperactivity disorder
ASD	Autism spectrum disorder
AtWBS	Atypical Williams-Beuren syndrome
BAZ1B	Bromodomain adjacent to zinc finger domain 1B
BDNF	Brain-derived neurotrophic factor
bFGF	basic fibroblast growth factor
bRGCs	Basal RGCs
CaMKII	Ca2+/Calmodulin-dependent protein kinase II
cAMP	Cyclic adenosine monophosphate
CCAT	Calcium channel-associated transcriptional regulator
cGMP	Cyclic guanosine monophosphate
CLIP2	CAP-GLY domain-containing linker protein 2
CNV	Copy number variation
D/P	Deletion of WBS syntenic region on mouse chromosome 5
DD	Distal deletion
DEGs	Differentially expressed genes
DNA	Deoxyribonucleic acid

DNS	Down syndrome
EB	Embryoid body
EGF	Epidermal growth factor
EIF4H	Eukaryotic initiation factor 4H
ELN	Elastin
EM	extracellular matrix
ER	Endoplasmic reticulum
ERK	Extracellular signal-regulated kinase
ES cell	Embryonic stem cell
FACS	Fluorescence-activated cell sorting
FISH	Fluorescence in situ hybridization
FBS	Fetal bovine serum
FDR	False discovery rate
FKBP6	FK506 binding protein-6
FOV	Fields of view
FPKM	Fragments Per Kilobase Of Exon Per Million Fragments Mapped
FZD9	Frizzled-9
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GO	Gene Ontology
GTF2I	General transcription factor 2I
GTF2IRD1	General transcription factor 2I repeat domain 1

GTF2IRD1/BEN/CREAM/GTF3	Alternate GTF2IRD1 symbols
IIH	Idiopathic infantile hypercalcemia
Inr	Initiator
IPA	Ingenuity Pathway Analysis
IPCs	intermediate progenitor cells
iPSC	Induced pluripotent stem cells
IQ	Intelligence quotient
KO	Knock-out
LAT2	Linker for activation of T cells-2
LCR	Low-copy repeat
LIMK1	Lin-11/Isl-1/Mec-3 kinase
MAPK	Mitogen-activated protein kinase
MECP2	Methyl CpG-binding protein 2
MEF	Mouse embrionic fibroblasts
mGluR	Metabotropic glutamate receptor
MLXIPL	MLX-interacting protein-like
MLPA	multiplex ligation dependent probe amplification
MSC	Mesenchymal stem cell
MRI	magnetic resonance imaging
mRNA	Messenger RNA
NCSC	Neural crest stem cells

NPC	Neural progenitor cells
NuFF	Newborn foreskin fibroblast cells
oRG	Outer radial glia
PD	Proximal deletion
PDGF	Platelet-derived growth factor
qPCR	Quantitative polymerase chain reaction
RGCs	Radial glial cells
RNA	Ribonucleic acid
SRF	Serum response factor
STX1A	Syntaxin 1A
SVAS	Supravalvular aortic stenosis
SVZ	subventricular zone
WBS	Williams-Beuren syndrome
WBSCR	Williams-Beuren syndrome critical region
WINAC	WSTF including the nucleosome assembly complex
WICH	WSTF-ISWI chromatin remodeling complex
WT	Wild type

Figures Index

1. Abstract

Symmetrical gene dosage imbalances at 7q11.23 chromosomal region cause two unique neurodevelopmental diseases, Williams Beuren Syndrome (WBS) and the 7q11.23 microduplication associated to autistic spectrum disorder (7dup-ASD). Although both these diseases share common features such as intellectual disability and craniofacial dysmorphism, they can be distinguished by distinct social and language abilities: WBS patients characterized by hypersociality and comparatively well-preserved language skills while 7dup-ASD is associated with impairment in social interaction and communicative skills. The involvement of same genetic interval in these disease, points out to small subset of dosage-sensitive genes affecting cognition, social behavior and communication skills.

Among the genes in the deleted region, some were shown to contribute to the abnormalities in these patients through transgenic mice models and individual case reports. However, the precise cellular and molecular phenotypes associated with these syndromes in disease-relevant cell-types are unknown due to the scarce availability of primary diseased tissues. Transcription factor induced somatic cell reprogramming has bypassed such fundamental limitation and has enabled us to model human diseases, elucidate their pathogenesis and discover new therapeutics by screening small chemicals/drugs on these models. During my PhD studies, I focused on the functional dissection of these complementary diseases at the level of transcriptional deregulation in patient-derived iPSC and its differentiated derivatives such as neural crest stem cells, mesenchymal stem cells, and neural progenitors. To this end, we have assembled a unique cohort of typical WBS, atypical WBS (patient with a partial deletion) and 7dup-ASD patients (along with unaffected relatives), and then I used mRNA reprogramming to establish and characterize at least 3 independent

iPSC lines from a total of 12 individuals. High throughput mRNA sequencing on iPSC revealed critical transcriptional derangements in disease-relevant pathways already at the pluripotent state. These alterations found to be selectively amplified upon differentiation into disease-relevant lineages, thereby establishing the value of large iPSC cohorts in the elucidation of disease-relevant developmental pathways. Finally, we created an open-access web-based platform to make accessible our multi-layered datasets and integrate contributions by the entire community working on the molecular dissection of the 7q11.23 syndromes.

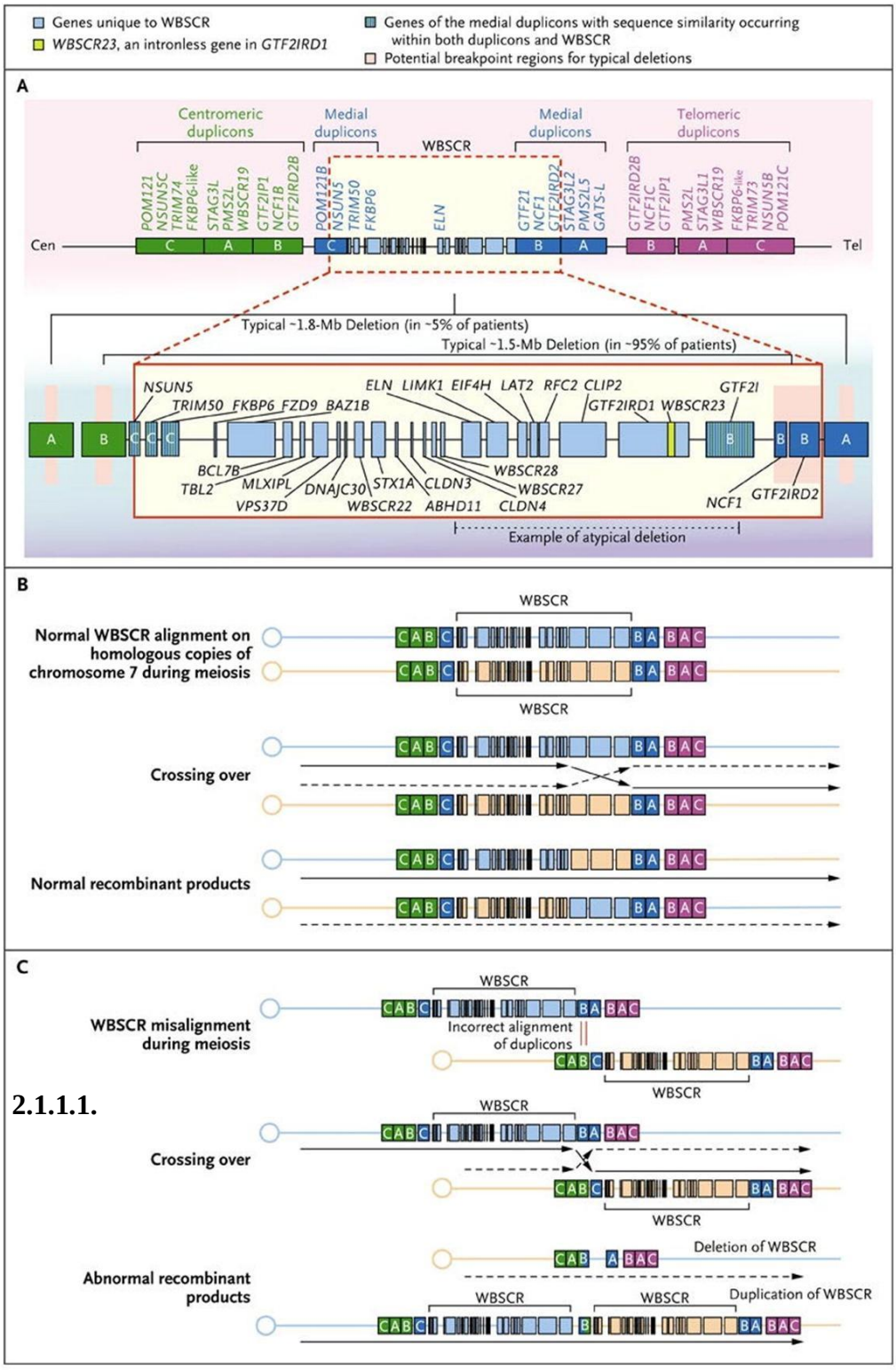
2. Introduction

The q11.23 genetic interval on human chromosome 7 contains genes that play crucial role in two human genetic neurodevelopmental diseases: Williams Beuren Syndrome (WBS; OMIM 194050) and Williams-Beuren region duplication syndrome (also known as Somerville-van der Aa syndrome, OMIM 609757) (Pober 2010, Sanders, Ercan-Sencicek et al. 2011). The hemizygous deletion of this region containing 26-28 genes (1.5-1.8 Mbp) causes WBS, a condition characterized by craniofacial dysmorphic features, cardiovascular defects and intellectual disability; further distinguished by hypersociality and comparatively well-preserved language skills (Pober 2010). On the contrary, the duplication of the same interval that has been recently reported in a subset of ASD patients (7dup-ASD) is associated with impairment in social interaction and communicative skills (Sanders, Ercan-Sencicek et al. 2011). This points out to the importance of small subset of dosage-sensitive genes within this interval affecting cognition, social behavior and communication skills.

2.1. Mechanisms of rearrangements at 7q11.23

The 7 q11.23 microdeletion/duplication occurs due to the unique genetic architecture of this interval. Williams‑Beuren syndrome chromosome region (WBSCR) is flanked by highly homologous clusters of genes/pseudogenes (also known as low-copy-repeat blocks or duplicons) (Pober 2010). The high degree of sequence homology among these flanking duplicons, as well as their proximity to each other, makes WBSCR highly susceptible to non-allelic homologous recombination (NAHR) that in turn mediates duplications and deletions during meiosis. Patients with deletion of WBSCR carry one copy of all genes within this

interval while patients with duplication carry three copies of this interval (Merla, Brunetti-Pierri et al. 2010, Pober 2010). The lists of the WBSCR genes and the mechanism of the rearrangement are depicted in Fig. 1.1.



2.1.1.1.

Fig. 1.1 The Williams‑Beuren syndrome chromosome region (WBSCR) on chromosome 7.

Panel A shows the list of WBS genes and that the WBSCR located between flanking duplicons. Panel B shows normal pairing of the two copies of the WBSCR during meiosis caused by alignment of the duplicons on the chromosome 7 homologues. Panel C shows abnormal pairing of the two copies of the WBSCR during meiosis, caused by misalignment of the duplicons due to their partial homology. Crossing over can result in abnormal recombinant products, either deletion of the WBSCR (causing Williams‑Beuren syndrome) or duplication of WBSCR (causing 7dup-ASD) (Adapted from (Pober 2010)).

2.2. Diagnosis

Historically WBS was diagnosed by FISH on metaphase chromosomes using a probe for the Elastin gene (Ewart, Morris et al. 1993, Merla, Brunetti-Pierri et al. 2010). In spite of being labor-intensive, time-consuming, and not highly accurate, FISH is yet the most wildly used method for clinical diagnosis of WBS. However, FISH interpretation for duplications makes interphase FISH highly problematic for the diagnosis of 7q11.23 duplication syndrome. For all these reasons, various techniques have been recently developed and exploited to improve the detection methods including qPCR, multiplex ligation dependent probe amplification (MLPA) and aCGH (Shaffer, Kennedy et al. 1997, Sellner and Taylor 2004, Schubert and Laccone 2006, Merla, Brunetti-Pierri et al. 2010, Sanders, Ercan-Sencicek et al. 2011).

In brief, qPCR is a precise method which allows estimation of the relative quantity of the analyzed locus assays within and outside the segmental aneuploidy (Schubert and Laccone 2006). Another efficient and reliable method is MLPA which can assess the dosage of the multiple genomic loci based on the synthetic probe (Sellner and Taylor 2004); and finally aCGH is a microarray based technology which can detect the copy number variations in highly sensitive way (Sanders, Ercan-Sencicek et al. 2011). Interestingly, thanks to these

methods, it has been found that there is a high frequency of parental transmission in 7q11.23 duplication patients which is in contrast with the rarity of parental transmission in the WBS (Merla, Brunetti-Pierri et al. 2010).

2.3.Williams Beuren syndrome

Williams Beuren syndrome (also known as William's syndrome) is a developmental disorder with a prevalence of roughly 1-10,000 person (Pober 2010). Although teratogenicity of vitamin D initially considered as the cause of disease based on the experiments that linked vitamin D with some symptoms of William's syndrome such as craniofacial and cardiovascular abnormalities, later studies showed that it's an autosomal dominant genetic disorder (Friedman and Roberts 1966, Morris, Thomas et al. 1993, Sadler, Robinson et al. 1993).

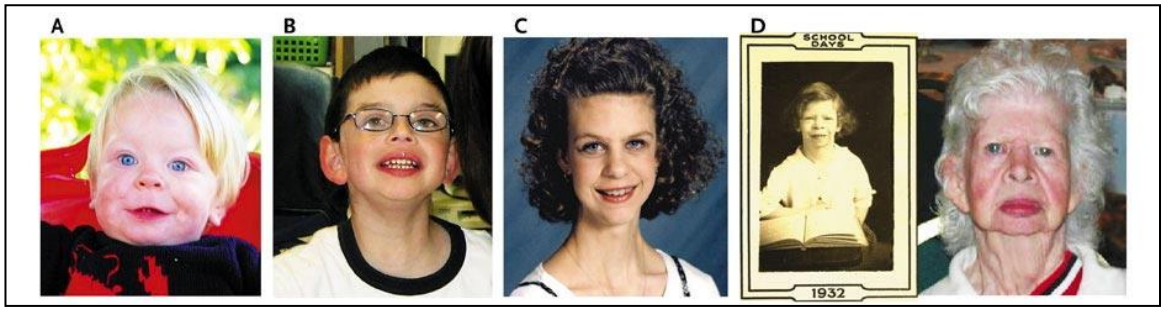


Fig.1.2 Patients with Williams-Beuren Syndrome.

Four unrelated patients with Williams-Beuren syndrome are shown in Panels A through D. The young child (Panel A) has a flat nose bridge, upturned tip of nose, long philtrum, mild periorbital puffiness, full cheeks, and a delicate chin. The school-age child (Panel B) has full lips, a wide mouth, and mildly increased interdental spacing. The young adult (Panel C) has a prominent nose and nasal tip, a wide mouth, and a full lower lip. Panel D shows a patient at 12 years of age (left) and at 83 years of age (right) (Adapted from (Pober 2010)).

As mentioned above, the hallmarks of the disease are low to mild intellectual disability, subtle to dramatic craniofacial features (Fig.1.2) and cardiovascular abnormalities which are listed and discussed below along with the possible role of individual genes in each phenotype:

2.3.1. Neurodevelopmental abnormalities

WBS patients have impaired (mild-to-moderate) intellectual disability with full-scale IQ averaging 50 to 60. They also exhibit dramatic weaknesses in visuospatial skills that is ability to visualize an object as a set of parts and construct a replica of the object from those parts. Fig. 1.3 compares the spatial deficits in William's syndrome with age and IQ match individual affected by Down syndrome (Bellugi, Lichtenberger et al. 1999, Meyer-Lindenberg, Mervis et al. 2006).

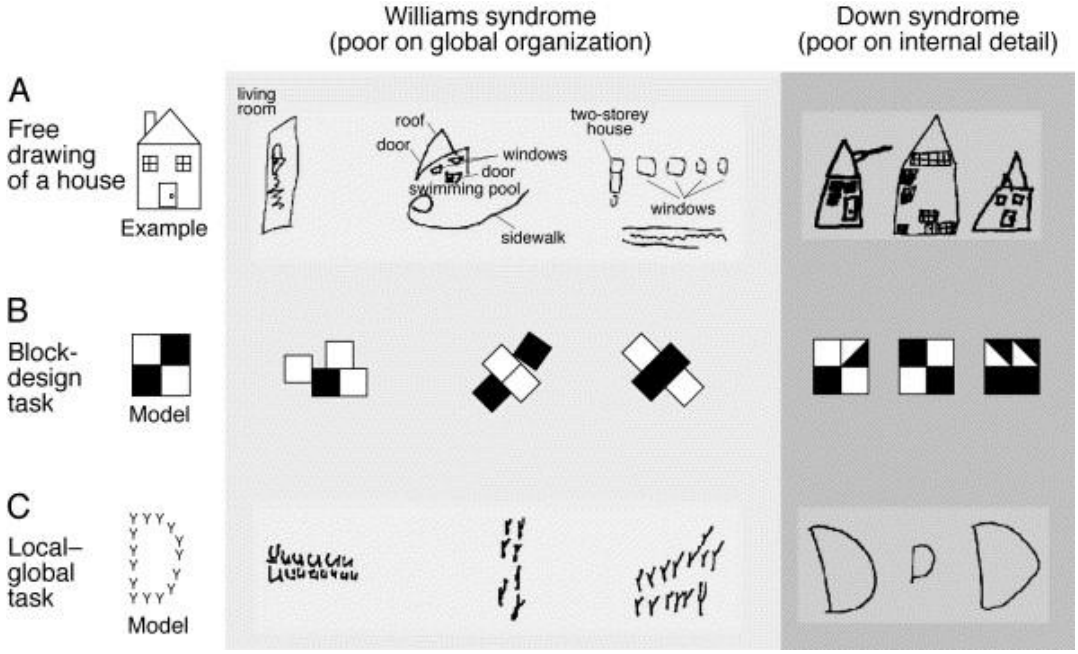


Fig. 1.3 Different spatial deficits in William's syndrome (WBS) and Down syndrome (DNS).

A) The drawings by adolescents and adults with WBS contain many parts of houses but they are not organized coherently. In contrast, the drawings of age-matched and Full Scale IQ-matched DNS adults are simplified but have the correct overall gestalt. B) In the block-design

task both subjects with WBS and subjects with DNS fail, but they fail in very different ways. C) In the Delis hierarchical processing task, subjects are asked to copy a large global figure made of smaller local forms (a 'D' made out of 'Y's). Again, both groups fail but in significantly different ways: subjects with WBS tend to produce the local elements sprinkled across the page, whereas age-matched and Full Scale IQ-matched subjects with DNS tend to produce only the global forms. (Adapted from (Bellugi, Lichtenberger et al. 1999)).

WBS patients also exhibit relative strengths in auditory rote memory, selected aspects of language, recognition and discrimination and social and interpersonal skills. Another well described feature of WBS patients is hypersociability which is combined with anxiety disorder, phobic disorder, attention deficit‐hyperactivity disorder in 50-90 percent of the WBS adults which over all, have a major impact on the quality of life of most people with Williams Beuren syndrome (Merla, Brunetti-Pierri et al. 2010, Pober 2010).

Neurologic examination and brain imaging using standard magnetic resonance imaging (MRI) reveals an overall 10 to 15% reduction in cerebral volume, with preserved cerebellar volume. Functional MRI studies also suggest that impaired limbic circuitry may underlie the unique anxiety profile of Williams–Beuren syndrome (Fig. 1.4) (Meyer-Lindenberg, Mervis et al. 2006, Jarvinen-Pasley, Bellugi et al. 2008).

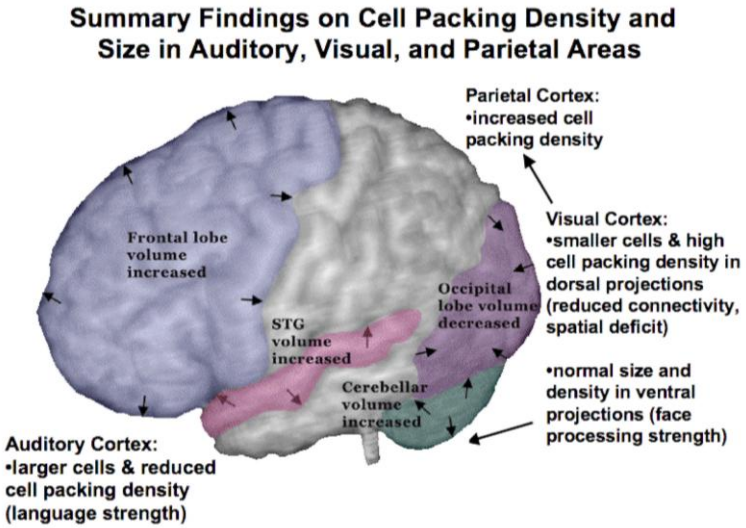


Fig. 1.4 Summary of neurobiological findings in WBS.

(Adapted from (Jarvinen-Pasley, Bellugi et al. 2008)).

Several genes have been implicated to be involved in the cognitive and behavioral phenotypes of WBS through mouse model studies:

CAP-Gly domain-containing linker protein 2 (CLIP2) belongs to a family of cytoplasmic linker proteins and regulate the cytoskeleton through the microtubule network. CLIP2 is abundantly expressed in neurons of the hippocampus, piriform cortex, and olfactory bulb. The mice model of this gene exhibit hippocampal dysfunction revealed by deficits in contextual fear conditioning and altered synaptic plasticity, however, no changes in the size or structure of the brain were observed. Both in hetrozygotes and homozygotes models impaired motor coordination on some tasks were observed but no changes in amygdale function or anxiety were seen (Hoogenraad, Koekkoek et al. 2002).

LIM kinase 1 (LIMK1) is a serine protein kinase involved in regulation of cytoskeletal integrity and remodeling through organization of the actin dynamics (Proschel, Blouin et al. 1995). LIMK1 is expressed in the central nervous system and in particular accumulates at the level of mature synapses, suggesting that it could play a role in synapse formation and/or maintenance (Scott and Olson 2007). In addition, the role of the actin remodeling in the establishment and modification of dendritic spines especially in the hippocampus suggest its putative role in the formation and maintenance of memory and learning (Nimchinsky, Sabatini et al. 2002). Likewise CLIP2, LIMK1 mice model shares some common phenotype with CTIP2 such as hippocampal dysfunction and no altered abnormality in brain size and structure. However, Limk1-null mice in contrast exhibited altered dendritic spine morphology

in pyramidal neurons which is also observed in many syndromes such as Down and fragile X syndromes which all share intellectual disability as a common phenotype. No analysis of the Limk1 heterozygotes was reported in the same study (Kaufmann and Moser 2000).

General transcription factor 2-I (GTF2I), GTF2I repeat domain containing protein 1 and 1 (GTF2IRD1) , and GTF2I repeat domain containing protein 2 (GTF2IRD2) all belong to the TFII-I gene family encoding transcription factors with multiple helix-loop-helix (HLH)-like domains, also known as I-repeats (Hinsley, Cunliffe et al. 2004). Genomic alignments suggest that GTF2IRD2 is a truncated version of GTF2I. GTF2IRD2 is deleted only in WBS patients with the rarer 1.84-Mb deletions, while GTF2IRD1 and GTF2I are deleted in all cases with typical deletions (Bayes, Magano et al. 2003).

Studies on the heterozygous and homozygous mice for Gtf2ird1indeed showed that both of these mice have some phenotypes which correlate with increased sociability and lack of social anxiety in WBS patients. For instance these mice showed increased social interaction, reduced anxiety and impaired amygdala-based learning and memory. However, unlike, CLIP2 and LIMK1 mice, hippocampal function showed to be intact. Enhanced serotonin receptor1A-mediated responses in layer V pyramidal neurons of the pre-frontal cortex were also reported in Gtf2ird1mice model (Palmer, Tay et al. 2007, Young, Lipina et al. 2008).

GTF2I is ubiquitously expressed and it can interact with multiple proteins and DNA, linking signal transduction to transcription. GTF2I also acts as a multifunctional transcription factor that can bind enhancer (E-box) and core promoter (Inr) elements in response to upstream signaling events (Roy, Du et al. 1997). The GTF2I genes have been also nominated initially as the candidate genes involved in cognitive and behavioral profile of WBS patients

through the study of patient with atypical deletion. Enkhmandakh et al. generated a Gtf2i gene-trap mouse model using a gene-trap cassette technology in which the cassette was inserted into intron 3 of Gtf2i. They reported that this mouse does not survive later than E10.5, and embryonic hemorrhage and cardiovascular malformations were discerned at E9.5. While about 60% of the homozygote embryos had neural tube defects, growth delay was apparent in the heterozygote mice (Enkhmandakh, Makeyev et al. 2009).

In another attempt Sakurai et al. (2010) generated a mouse model using a similar approach. Similar to the previous study this mouse did not produce any homozygouse animals and Gtf2i-null embryos were exencephalic however, they did not exhibit developmental and cardiovascular abnormalities which are in contrast with former study. In terms of behavioral assays, Gtf2i+/- animals did not exhibit any alterations in spatial and non-spatial learning and memory, anxiety and neuromotor function (Sakurai, Dorr et al. 2011).

2.3.2. Cardiovascular abnormalities

In general cardiovascular defects are the major cause of death in WBS patients. They occur 25-100 times more with respect to the healthy individuals. Supravalvular aortic stenosis (SVAS) is the most common cardiovascular abnormality of Williams–Beuren syndrome affecting approximately 70 percent of the patients (Pober 2010).

Previous studies have found a strong link between Elastin (ELN) and SVAS and other connective tissue abnormalities of WBS patients (Ewart, Morris et al. 1993). Various studies also showed that SVAS is due to the over growth of the smooth muscle cells within vascular media (Pober 2010). Elastin is mainly responsible for the formation of the elastic fibers of the extracellular matrix throughout connective tissue of the body. Reduction of vascular

elasticity, due to elastin haploinsufciency, may increase the hemodynamic stress to the endothelium, which in turn leads to over growth of smooth muscle and subsequently narrowing of the vessels (Karnik, Brooke et al. 2003). However, the pathogenesis of the arteriopathy in WBS may be more complex, and is possibly related to other genes in the WBS critical region such as NCF1 (Del Campo, Antonell et al. 2006, Merla, Brunetti-Pierri et al. 2010).

Homozygous knock-out mice for elastin is not viable and die right after birth due to aortic obstruction caused by smooth muscle cell proliferation in the arterial wall. However, heterozygous mice are viable and produce half of the Eln mRNA and protein. They also exhibit hypertension, thinner elastic lamellae, decreased aortic compliance, and mild cardiac hypertrophy (Li, Brooke et al. 1998, Li, Faury et al. 1998). Importantly, the Eln mouse model does not show SVAS but interestingly, Eln heterozygosity in a transgenic mouse carrying a human version of ELN on a bacterial artificial chromosome resulted in thickening of the wall of the ascending aorta which suggests a fundamental difference in the function of the mouse and human ELN gene in the developing aorta (Hirano, Knutsen et al. 2007) .

Another gene which has been shown to have a role in the development of heart is BAZ1B (encoding WSTF) that has roles in DNA repair, replication, transcriptional activation and repression, and also possesses histone H2A kinase activity (Barnett and Krebs 2011). WSTF is a shared subunit of three distinct chromatin remodeling complexes; WICH (WSTF-ISWI chromatin remodeling complex) for DNA repair and WINAC (WSTF including the nucleosome assembly complex) for transcriptional control and B-WICH (Barnett and Krebs 2011, Barnett, Yazgan et al. 2012). BAZ1B knockout mice exhibit major heart defects and dies shortly after birth, however, heterozygous mice shows range of developmental heart

abnormalities similar to those have been reported in WBS. Moreover, expression of some key transcriptional regulators involved in heart development such as Gja5 and Irx3 shown to be altered in these mice (Yoshimura, Kitagawa et al. 2009).

2.3.3. Craniofacial dysmorphisim

Craniofacial abnormalities considered as one of the distinctive features of these patient and they include (but are not limited to) broad forehead, periorbital fullness, epicanthal folds, flat nasal bridge, a short upturned nose, long philtrum, and wide mouth with full lips, full cheeks, small jaw excessive and interdental spacing (Pober 2010).

The two important genes proposed so far to be involved in craniofacial feature of WBS patient based on the existing mouse models are Gtf2ird1 and Baz1b. Briefly Gtf2ird1 homozygous knock out mouse exhibit periorbital fullness and a short snout misaligned jaws and a twisted snout. However, the heterozygous mice, showed normal craniofacial development (Tassabehji, Hammond et al. 2005, Palmer, Tay et al. 2007, Young, Lipina et al. 2008) . Baz1b knock-out mice had significantly different skulls mainly as a results of reduction in the parietal and nasal bones as well as a relative hypoplasia of the lower jaw while the heterozygote mice had narrower and shorter craniums in comparison to the wild-type mice along with reduced size of the posterior region of the lower jaw. Importantly, Baz1b expression is strongly up-regulated in all the major facial primordia from early in embryogenesis including the cranial neural crest-derived mesenchyme that drives facial morphogenesis. These results also suggest that BAZ1B could be important for cranial development (Ashe, Morgan et al. 2008, Yoshimura, Kitagawa et al. 2009).

Having said that, the most striking craniofacial abnormalities such as shorter cranial base and narrowing of the posterior part of the skull observed in mice lacking human proximal region spanning Trim50-Limk1. Importantly, no differences in cranial morphology of mice lacking the human distal region (Limk1–Gtf2i) were seen which is in contrast with the fact that in atypical WBS with distal deletion craniofacial features are more evident implying to the complexity and difference of craniofacial abnormalities in mice and human (Li, Roy et al. 2009).

All together, these studies further highlight the role of BAZ1B as a tantalizing candidate since both in heterozygous knockout mice and in the context of larger deletion the phenotype were evident.

2.3.4. Calcium abnormalities

Ranging from severer (during infancy) to mild (in general), hypercalcemia can affect 5-50 percent of the WBS patients. There is no confirmed mechanism for hypercalcemia in these patient however, various mechanism such as vitamin D sensitivity, increased 1,25-dihydroxyvitamin D levels, and defective calcitonin synthesis or release have been suggested so far (Pober 2010). In line with the transient hypercalcemia seen in infants and children with WBS, heterozygous Baz1b mice showed elevated serum calcium levels (Yoshimura, Kitagawa et al. 2009).

2.3.5. Diabetes Mellitus

In the recent studies it has emerged that majority of WBS patients suffer from diabetes Mellitus and that there is a high prevalence of impaired glucose tolerance among patients with Williams Beuren syndrome (Pober 2010).

Through the studies of the knockout mice models so far two genes has been implicated to play a role in glucose dysregulation. First, syntaxin 1A was shown to have a role in membrane vesicle fusion and pancreatic cell exocytosis of insulin granules, and more interestingly mouse models with altered STX1A expression levels showed significant alteration in glucose metabolism due to abnormal insulin secretion, nominating this gene as a important candidate for the observed phenotype in these patients. (Lam, Leung et al. 2005, Ohara-Imaizumi, Fujiwara et al. 2007).

The second candidate which could be involved in metabolism of glucose in WBS patient is MLXIPL which is an element binding protein and regulate the expression of a liver enzyme responsible for the carbohydrate response synthesis of fatty acids and triglycerides (Iizuka and Horikawa 2008). The elevated plasma glucose level in Mlxpl-null mice is in line with the phenotype observed in WBS patients, however, in contrast to WBS patients, Mlxipl-null mice have reduced fatty acid synthesis and lowered fat deposition (Osborne 2010).

2.4.7q11.23 duplication syndrome

Although from the time the genetic basis for WS was discovered by Ewart et al., in 1993, there was an assumption that there also could be a syndrome caused by a duplication of the WBS genetic interval, 7q11.23 duplication syndrome has emerged only recently and therefore the full clinical spectrum has not been discussed extensively (Ewart, Morris et al. 1993, Somerville, Mervis et al. 2005). One reason for the fact that this syndrome remains often undiagnosed, could be since patients with duplications often have milder pathological consequences than the reciprocal deletions. Beside milder facial dimorphic abnormalities, the most striking features of 7q11.23 duplication is impairment in social interaction and communicative skills similar to autistic spectrum disorders and indeed there is an increased

prevalence of autism among these patients (Velleman and Mervis 2011). The involvement of same genetic interval in these two 7q11.23 associated diseases, points out to small subset of dosage-sensitive genes affecting cognition, social behavior and communication skills.

Fig.1.5 Three children who have 7q11.23 duplication syndrome, aged 2 years, 7 years, and 9 years.

(Adapted from (Velleman and Mervis 2011))

2.4.1. Facial characteristics

So far facial dysmorphic features have been reported in all of 7q11.23 duplication cases. These craniofacial features include a broad forehead, high broad nose, neatly placed straight eyebrows. Moreover, short philtrum and thin lips are two distinctive features which are opposite to what observed in WBS patients (Fig 1.5). In addition, some authors have commented on a slight asymmetry of the face and macrocephaly that has been reported in a some cases (Somerville, Mervis et al. 2005, Berg, Brunetti-Pierri et al. 2007, Torniero, Dalla Bernardina et al. 2008, Van der Aa, Rooms et al. 2009, Merla, Brunetti-Pierri et al. 2010, Velleman and Mervis 2011)

2.4.2. Cardiovascular abnormalities and connective tissue involvement

Cardiovascular abnormalities are not frequently reported and only affect around 20 percent of people with 7q11.23 duplication. Beside ventricular and atrial septal defect, the most common problem so far was patent ductus arteriosus which is a failure in the closure of the channel between the aorta and the pulmonary artery that takes blood to the lungs (Merla, Brunetti-Pierri et al. 2010).

2.4.3. Neurodevelopmental abnormalities

Regarding the neurological problems, epilepsy (with the prevalence of 20 percent) and hypotonia are the most commonly reported ones. In line with the WBS patient, the brain MRI abnormalities have frequently been reported, however no consistent brain abnormalities have been reported so far (Torniero, Dalla Bernardina et al. 2008, Van der Aa, Rooms et al. 2009).

In term of cognitive and behavioral profile, the majority of the patients are developmentally delayed with few exceptions with normal IQ (Van der Aa, Rooms et al. 2009).

Either expressive and/or receptive language delay has been seen in almost all patients. Such language impairment along with the spared of visuospatial cognitive skill in patient with 7q11.23 duplication is in contrast with typical cognitive profile of the WBS patient. As discussed previously, WBS patients exhibit a relative strength in verbal skill and are severely impaired in visuospatial skills (Torniero, Dalla Bernardina et al. 2008, Pober 2010).

Another important feature that has been noted in 7q11.23 duplication patients is deficits in social interaction including poor eye contact, poor social interaction, limited facial

expressions, repetitive behaviors and repetitive speech which are all considered as formal diagnosis of autism or autism spectrum disorder (Berg, Brunetti-Pierri et al. 2007, Van der Aa, Rooms et al. 2009).

Recently, Osborne group has generated mice with varying numbers of Gtf2i copies. Interestingly, comparing to the mouse pups which are harboring one or two copies of GTF2i, pups with extra GTF2i copies exhibited increased maternal separation-induced anxiety (Mervis, Dida et al. 2012).

2.5. Rationale and aims of the study

Despite of such progress in understanding of the physiopathology of the 7q11.23 associated disease, the attempts to correlate genotype with phenotype in WBS have been only moderately successful as they mainly relied on:

2.5.1. Individuals with atypical deletions

The identification of individuals with atypical deletions of the critical region in which only a subset of genes are hemideleted can be of immense value. However, these individuals are extremely rare and so far less than fifty individuals with such deletion have been identified. Moreover, the exact breakpoints of each deletion have not been established in many cases. In addition, the third major problem in this regard is that, these patients have not been evaluated by the same physicians and thus they have not been subjected to the same battery of clinical, cognitive, and psychological testing. More importantly, there's too much variability in phenotype between WBS patients (suggesting a high modulation of the phenotype due to the environment or remaining genetic background) to draw a conclusion from a single or a few case reports. (Merla, Brunetti-Pierri et al. 2010).

2.5.2. Generation of mouse models

Several mouse models knocked-out for single WBS genes have been generated in order to understand the molecular basis of WBS. Out of the 26 genes deleted in WBS, so far, 11 published mouse models exist most of which were generated through conventional gene targeting techniques in embryonic stem cells based on homologous recombination (Osborne 2010).

Although mouse models of WBS provide invaluable insight into the role of both individual and combinatorial gene disruption over a wide spectrum of analyses, however, they still cannot fully recapitulate human complex traits such as neurocognitive features. This is due to the fact that the primate cortex, and particularly the human cerebral cortex is different in several aspects from the rodents. For examples, there is a marked increase in the size of the cerebral cortex relative to the rest of the nervous system which could be the results of complexity and diversity of its developing stem cell populations (Shi, Kirwan et al. 2012). Regarding the musculoskeletal traits, mice are obviously quadripedal and thus these features will present differently than they might be in bipedal humans. Beside higher metabolic rate, earlier reproductive age and a far shorter lifespan than humans, mice have adapted to environments, predators and pathogens which are not true in the case of humans. These all could justify why it was anticipated that at least 20% of human important genes could be non-essential in the mouse, which means that homozygously deletion in those genes will not result in lethality. And finally, comparing mice and human from genetic point of view there are 300 genes which are unique to each species, making it impossible to study them in other organism (Osborne 2010).

2.5.3. Cells and tissues from WBS individuals

In terms of gauging WBS pathogenesis in human cells, so far two genome-wide transcriptomic analyses were performed, respectively on EBV-immortalized lymphoblasts or fibroblasts from WBS patients (Antonell, Vilardell et al. 2010, Henrichsen, Csardi et al. 2011). Our lab has performed a meta-analysis of these results starting from the re-evaluation of raw data, and found virtually no overlap between the sets of genes that were significantly deregulated in each study between WBS samples and controls, pointing to the acute need to generate meaningful, disease-relevant cell types from WBS patients (Fig1.6).

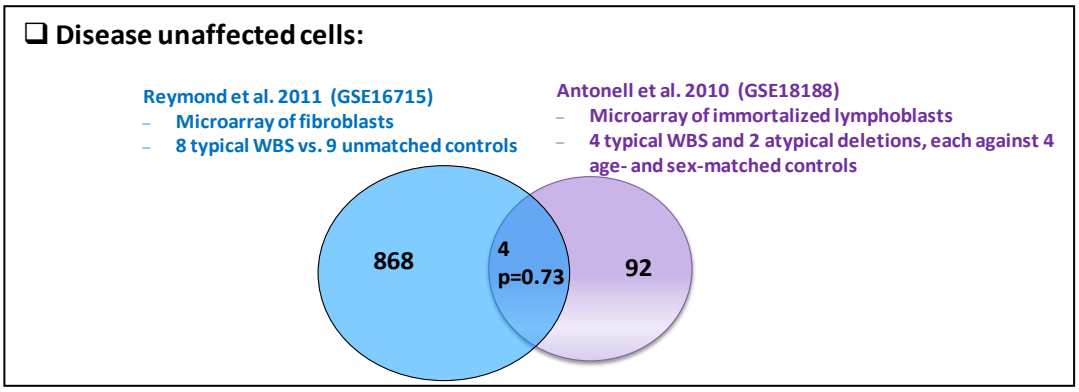


Fig. 1.6 The Venn diagram showing the overlap of the differentially expressed genes from two transcriptomic data;

fibroblast and immortalized lymphoblasts coming from WBS patient and healthy individuals. Bioinformatic analysis identifies no significant overlap between published transcriptomic studies pointing to the acute need for pathophysiologically meaningful cellular model.

2.5.4. iPSC-based technology as a platform for disease modeling

Similar to any other scientific achievements, induced pluripotent stem cell technology was established on the basis of the various fundamental findings some of which have been summarized in Fig. 1.7. Combination of all three streams of research enabled Shinya Yamanaka to design the experiments that lead to the identification of the factors which can

convent somatic cell to embryonic stem like state (Takahashi, Tanabe et al. 2007, Yamanaka 2012). Briefly the generation of the tadpoles from unfertilized eggs that had received a nucleus from the intestinal cells of adult frogs reported in 1962 by John Gurdon could be considered as the first stream (Gurdon 1962).This finding were further consolidated by the generation of the first mammal through somatic cloning of mammary epithelial cells by Ian Wilmut and colleagues (Wilmut, Schnieke et al. 1997).

The second stream was two seminal studies by two groups in the same year showing the role of the master transcription factors in inducing the formation of the legs and myocytes in Drosophila and fibroblast, respectively (Davis, Weintraub et al. 1987, Schneuwly, Klemenz et al. 1987) and finally, the third stream was the generation of mouse ESCs in 1981(Evans and Kaufman 1981, Martin 1981), further followed by establishment of the culture conditions that enabled the long-term maintenance of pluripotency by Austin Smith (Smith, Heath et al. 1988) and optimal culture conditions with basic fibroblast growth factor (bFGF) by James Thomson (Thomson, Itskovitz-Eldor et al. 1998).

The first isolation of human embryonic stem cell by Thomson and colleagues in 1998 provided the opportunity of deriving large quantities of differentiated cell types using defined growth factors (Thomson, Itskovitz-Eldor et al. 1998). Such a breakthrough was considered as an improved and reliable physiological models to study human developmental biology and drug discovery. Over the past decade various studies conducted to demonstrate the feasibility of using human pluripotent stem cells (PSCs) for these applications(Grskovic, Javaherian et al. 2011). In spite of considerable advantages of ESC over routinely used immortalized or primary cells, ESC studies have faced to various issues. First, human ESC generated considerable ethical debates. Second only monogenic disorders could be studied as multifacto-

rial disease cannot be screened through pre-implantation diagnostics (Grskovic, Javaherian et al. 2011).

Indeed, only nine years after Thomson's work, the ground breaking study of Shinya Yamanaka demonstrated that human fibroblast cells can be converted to cells closely resembling ESC, so called induced pluripotent stem cells (iPSC) only through ectopic over - expression of four transcription factors namely; OCT4 (also known as POU5F1), SOX2, KLF4 and MYC. The iPSC cells were not only similar to ESC in terms of morphology, growth characteristics but also were able to differentiate and form embryoid bodies in vitro and teratomas in vivo. More importantly, upon injecting iPSC into blastocysts they were capable to generate the entire animal showing that that iPSC similar to ESCs, possess full developmental potential (Takahashi, Tanabe et al. 2007).

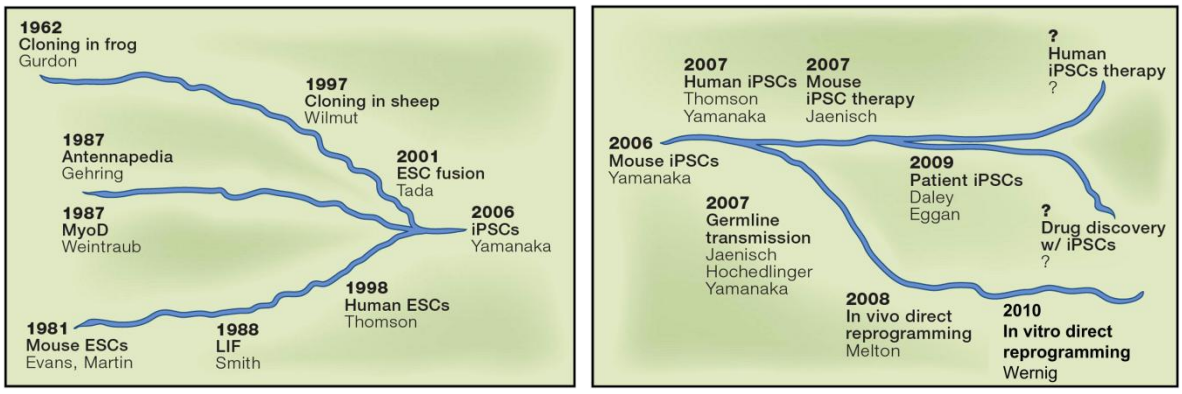


Fig. 1.7 Scientific streams led to or emerged from iPSC technology.

A) Three scientific streams that led to the development of iPSCs. B) New scientific streams that emerged from the development of iPSCs. (Adapted from (Yamanaka 2012)).

Soon after the initial report of mouse iPSCs, many group recapitulated the factor-based reprogramming in various systems and therefore new scientific streams have emerged from iPSC including cell replacement therapy, disease modeling and drug screening (Fig. 1.7)(Yamanaka 2012).

iPSC based disease modeling has enhanced the study of disease mechanisms and therapies. Following the first report of human iPSCs in which came out a yeat afte mice report, the field of disease modeling based on iPSC technology was initiated by the generation of iPSCs using somatic cells derived from aged patients (Dimos, Rodolfa et al. 2008) and subsequently many other diseases were modeled so far (Park, Arora et al. 2008), (Grskovic, Javaherian et al. 2011, Inoue, Nagata et al. 2014).

Identification of a disease-relevant cellular pathology is main goal in modeling any disease with iPSCs (Fig. 1.8). So far, modeling early onset diseases with strong genetic component and those with affect a highly defined cell or tissue type were the most compelling demonstrations (Ebert, Yu et al. 2009, Lee, Papapetrou et al. 2009, Marchetto, Carromeu et al. 2010, Brennand, Simone et al. 2011, Nguyen, Byers et al. 2011, Consortium 2012). In addition, for some disease with a known molecular mechanism such as spinal muscular atrophy or familial dysautonomia iPSC based assays which can capture relevant pathological mechanisms of the disease have been developed and further subjected for small-molecule screening purposes in order to identify of new potential compounds (Ebert, Yu et al. 2009, Lee, Papapetrou et al. 2009, Grskovic, Javaherian et al. 2011). However, modeling more genetically complex disease such as sporadic Alzheimer's disease or Parkinson's disease still requires further studies for the identification of the cellular phenotypes that relate to known aspects of disease pathology (Grskovic, Javaherian et al. 2011).

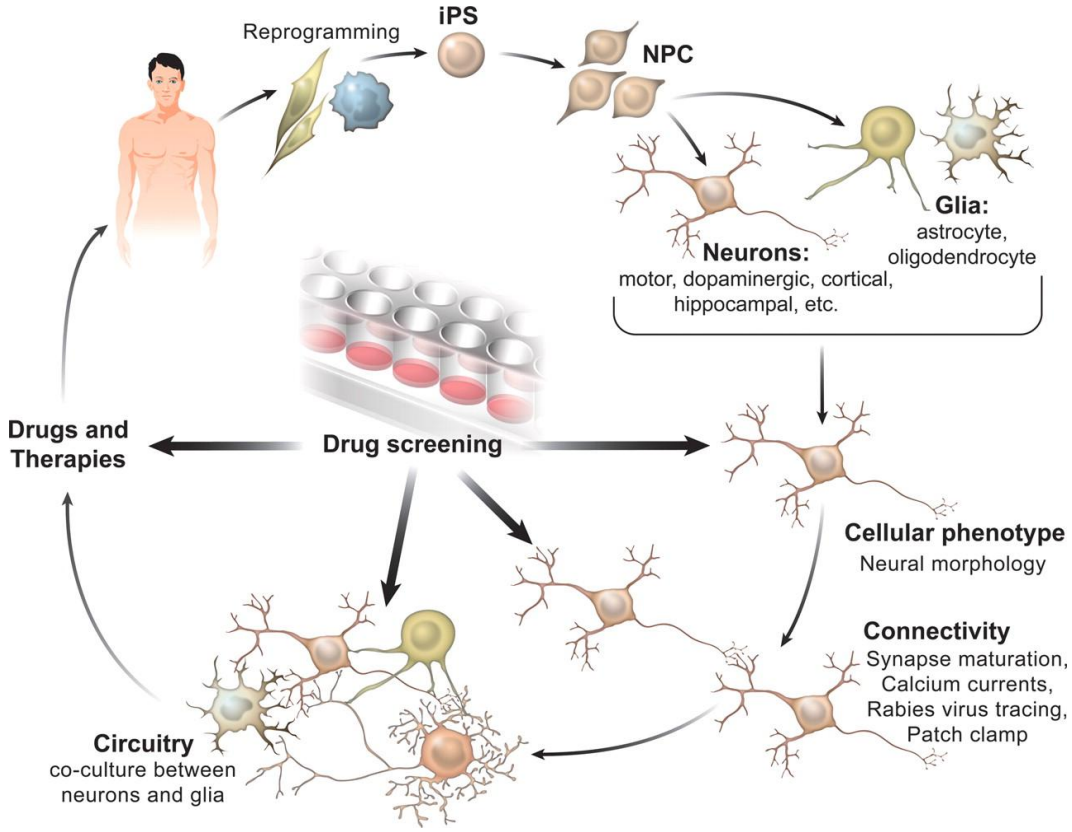


Fig. 1.8 iPSC to model neurodegenerative and neurodevelopmental diseases.

Human iPSC from neurologic patients and controls are generated after somatic tissue reprogramming (e.g. skin or blood cells). Neural progenitor cells (NPC) are generated and are further differentiated into neurons and/or glial cells. Neurons are then differentiated into subtypes of neurons such as dopaminergic, cholinergic, etc. Cellular phenotype is assessed by measuring neuronal morphology (i.e. process branching, spine density/size/maturation). Next, connectivity and circuitry integration can be analyzed by calcium influx transients, electrophysiology and transneuronal tracing with the rabies virus. In addition, the cross-talk between neurons and glia can be studied to tease out autonomous and non-autonomous aspects of the disease. Once a distinct disease-related phenotype is identified, drug-screening platforms can be developed to test compounds that improve cellular phenotype. Therapeutic compounds could emerge from the screenings, potentially benefiting neurologic patients (Adapted from (Marchetto, Winner et al. 2010)).

To date, several disease phenotypes has been reported from patient-derived cells which have been used to recapitulate in vitro models of various disease Some of the key examples of neurological disease are discussed below:

2.5.4.1.Spinal Muscular Atrophy

Spinal Muscular Atrophy (SMA) is a neuromuscular childhood disease with an autosomal recessive inheritance pattern. SMA is the most common cause of death among heritable disease in infants and unfortunately there are no effective treatments available to date. A mutations in the SMN1 gene results in a decrease in the levels of the survival of motor neuron (SMN) protein which subsequently leads to motor neuron degeneration. Screening in patient derived irrelevant cells such as fibroblast has not yielded in any compound which can enter to the clinic probably due to the fact that the mechanisms that regulate SMN protein levels in non disease affected cells are substantially different from those in human motor neurons in vivo(Crawford and Pardo 1996, Grskovic, Javaherian et al. 2011).

Ebert and colleagues generated two iPSC lines from a patient with SMA and an unaffected relative and differentiated them into motor neurons. Importantly, number of motor neurons was shown to be reduced in the cells derived from patients with SMA. This key experiment demonstrated for the first time that reprogramming and differentiation process can faithfully capture and recapitulate the disease phenotype. Although two compounds (valproic acid and tobramycin), were shown to increase the number of SMN-rich structures in the patient-derived iPSC, it remains to be seen whether these compounds can exhibit the same effect in motor neurons and thus rescue motor neuron loss in patients (Ebert, Yu et al. 2009, Grskovic, Javaherian et al. 2011).

2.5.4.2.Familial dysautonomia

Familial dysautonomia is a disorder which affects the survival of sensory, sympathetic and parasympathetic neurons as a result of abnormal migration of neural crest cells. Like SMA, there are not any effective treatments or appropriate disease models for this disease. In majority of the patients it is caused by mutations in the gene encoding IκB kinase complex-associated protein (IKBKAP) resulting in the skipping of exon 20 (Slaugenhaupt, Blumenfeld et al. 2001).

Lee, Studer and colleagues derived iPSCs from three young patients with familial dysautonomia and subsequently differentiated those lines into neural crest cells. They found a defect in the migration and neurogenesis in the neural crest cells. Through an screening approach they found that kinetin can markedly reduce the splicing and neurogenesis defect in iPSC-derived neural crest cells of patients with familial dysautonomia. This study further confirmed the importance of iPSC-based cellular models for phenotypic screening and drug discovery (Lee, Papapetrou et al. 2009, Grskovic, Javaherian et al. 2011).

2.5.4.3. Rett syndrome

In another effort, Marchetto et al. focused on Rett syndrome, a developmental neurological disease which is part of the larger group of autism spectrum disorders. This syndrome is caused by mutations in methyl-CpG-binding protein, a protein involved in DNA methylation(Chahrour, Jung et al. 2008). iPSCs generated from healthy controls and patients with Rett syndrome and then were differentiated into glutamatergic and GABergic neurons. By assessing neurogenesis, synapse number and neuronal morphology the authors did not observe any changes in neurogenesis. However, they were able to measure a substantial

reduction in synapse number as well as a reduction in the number of spines which were in line with previous reports observed in the post-mortem brains of patients with Rett syndrome. Further experiment using electrophysiology and calcium imaging showed that calcium oscillations and the frequency of spontaneous postsynaptic currents were decreased in the neurons of patients with Rett syndrome (Marchetto, Carromeu et al. 2010, Grskovic, Javaherian et al. 2011).

2.5.4.4.Parkinson's disease

Parkinson's disease (PD) is caused by the progressive loss of midbrain dopaminergic neurons and it is considered as the second most common neurodegenerative disorder in the world. However, there is not any known cure for Parkinson's disease, and neurodegeneration progresses leads to the worsening of symptoms and a loss of therapeutic efficacy. In two different attempts, patient-derived neurons from familial forms of Parkinson's disease have been exploited to model disease in vitro (Grskovic, Javaherian et al. 2011, Nguyen, Byers et al. 2011, Seibler, Graziotto et al. 2011).

In the first study Seibler and colleagues derived dopaminergic neurons from patients with mutations in the gene encoding PTEN-induced putative kinase 1, which is thought to regulate a protein that is associated with familial Parkinson's disease, parkin. Patient-derived dopaminergic neurons exhibited impaired recruitment of parkin to mitochondria, increased mitochondria copy number and increased expression of the mitochondrial regulator peroxisome proliferator-activated receptor-γ co-activator 1α. Strikingly, authers showed that these phenotypes could be rescued by exogenous expression of wild-type PTEN-induced putative kinase 1 (Grskovic, Javaherian et al. 2011, Seibler, Graziotto et al. 2011).

In the second attempt Nguyen and colleagues generated iPSCs from a patient with the most common mutations familial Parkinson's disease encoding leucine-rich repeat kinase. Patient-derived dopaminergic neurons showed increased vulnerability to stress by hydrogen peroxide, 6-hydroxydopamine and the proteasome inhibitor MG-132, in line with the previous findings that both genetic and environmental factors contribute to the development of Parkinson's disease(Nguyen, Byers et al. 2011).

2.5.4.5.Huntington's disease

Huntington's disease (HD) is a neurodegenerative disorder with an autosomal dominant inheritance pattern. HD patients suffer from progressive motor dysfunction, cognitive decline, and psychological problems. HD caused by an expansion of polyglutamine(CAG) repeats in the huntingtin (HTT) protein and the age of onset for these symptoms is correlated with the number of repeats, where more than 36 repeats is considered a pathological threshold(An, Zhang et al. 2012).

The HD Consortium reported a unique and well-characterized resource to elucidate disease mechanisms in HD through the generation and characterization of a large cohort of iPSC lines from HD patients and healthy controls. Importantly, microarray profiling revealed CAG-repeat-expansion-associated gene expression patterns that distinguish patient lines from controls, and early onset versus late onset HD. iPSC derived neural cells exhibited disease-associated alterations in electrophysiology, metabolism, cell adhesion, and ultimately cell death for lines with both medium and longer CAG repeat expansions. They also showed that lines harboring the longer repeats are the most vulnerable to cellular stressors and BDNF withdrawal, as assessed by a range of assays across consortium laboratories. The HD iPSC

collection provides a human stem cell platform for screening novel the therapeutics(Consortium 2012).

2.5.4.6.Schizophrenia

Schizophrenia (SCZD) is a genetic neurological disorder with a world-wide prevalence of 1%. The estimated inheritability of the disease is a 80–85%. Post-mortem studies have revealed some abnormalities such as reduced brain volume, cell size, spine density and abnormal neural distribution in the prefrontal cortex and hippocampus of SCZD brain tissue(Wong and Van Tol 2003). Although, the neuropharmacological studies have implicated dopaminergic, glutamatergic and GABAergic as the main cell types affected in SCZD, the molecular mechanisms underlying the disease state remain unclear. The first utility of iPSC for such a complex genetic psychiatric disorder were reported by Brennand et al. where they probed hiPSC based neuronal phenotypes and gene expression changes associated with SCZD. .They reprogrammed fibroblasts from SCZD patients into human induced pluripotent stem cells and subsequently differentiated these patient derived hiPSCs into neurons. In brief, SCZD hiPSC neurons exhibited diminished neuronal connectivity in conjunction with decreased neuritis number, PSD95-protein levels and glutamate receptor expression. Strikingly, following treatment of SCZD neurons with the antipsychotic loxapine, the main cellular and molecular elements of the SCZD phenotype were rescued (Brennand, Simone et al. 2011).

2.5.4.7.Timothy syndrome

Of the key examples of the iPSC based disease modeling for monogenic neurodevelopmental disorders were reported by Dolmetsch and his colleagues by modeling

Timothy syndrome which is caused by a missense mutation in the L-type calcium channel Cav1.2 that is associated with developmental delay and autism. They successfully generated cortical neuronal precursor cells and terminally differentiated neurons from induced pluripotent stem cells derived from patients with Timothy syndrome. Authors revealed that cells from these individuals have defects in calcium (Ca2+) signaling and activity-dependent gene expression, as well as abnormalities in differentiation, including decreased expression of genes that are expressed in lower cortical layers. Timothy syndrome derived neurons showed abnormal expression of tyrosine hydroxylase and increased production of norepinephrine and dopamine. Importantly, they could rescue the phenotype by treatment with roscovitine, a cyclin-dependent kinase inhibitor and atypical L-type–channel blocker2–4. These findings contributed significantly to understand how Cav1.2 regulates the differentiation of cortical neurons in humans and offered important insights into the causes of autism in individuals with Timothy syndrome (Pasca, Portmann et al. 2011). Despite of such a great reports, many issues regarding the utility and application of the iPSC cells has yet to be further elucidated some of which are discussed below:

2.5.4.8. Safe, efficient and reliable protocol for generation of iPSC

The historical and the most widely practiced method for generation of iPS is transduction of reprogramming factors via retro/lenti viruses. However, these methods are limited by: i) low efficiency of iPSC derivation; ii) the risk of insertional mutagenesis and iii) residual transgene expression from integrated vectors which may inadvertently affect the differentiation of iPSCs (Takahashi, Tanabe et al. 2007, Yu, Vodyanik et al. 2007, Warren, Manos et al. 2010).

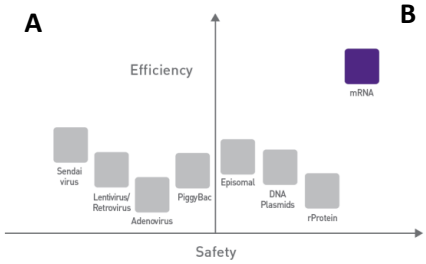
To tackle these issues, numerous non-integrating platforms based on DNA, RNA, miRNAs, and proteins have been developed to generate integration-free iPSCs, and the advantages and drawbacks have been extensively discussed previously (Gonzalez, Boue et al. 2011). Among these, RNA-based iPSC approaches using Sendai virus (Fusaki, Ban et al. 2009), miRNAs (Anokye-Danso, Trivedi et al. 2012), and mRNA transfection (Warren, Manos et al. 2010) avoid integration associated problems and therefore appear safer methods for future clinical applications.

Infection by Sendai virus has been recently used for the expression of pluripotency factors. Since it is a negative-sense, single-stranded RNA virus and does not go through a DNA intermediate, it can be considered as a highly efficient and integration free method for reprogramming (Fusaki, Ban et al. 2009, Ban, Nishishita et al. 2011). However, the main drawback of this method is that Sendai virus replicates in iPSC clones after reprogramming, therefore this approach requires a selection step followed by several passaging steps from the single-cell level to isolate virus-free iPSCs. Recently a temperature-sensitive mutant of Sendai virus has been introduced as a successful alternative method to remove the virus, though this method still requires a higher biosafety due to production of infectious virus particles (Ban, Nishishita et al. 2011, Yoshioka, Gros et al. 2013).

In a more promising attempt, Warren et al. first described the ability to reprogram human cells using modified mRNA with conversion efficiencies and kinetics superior to DNA-based methods. This approach involves the daily transfection of five individual mRNAs (OCT4, SOX2, KLF4, CMYC, and LIN28) over the 14-days reprogramming period. This technology was a major leap forward to safely and effectively reprogram human cells (Warren, Manos et al. 2010).

The impact of specific microRNAs on the promotion of induced pluripotency had previously been demonstrated. While the work of Anokye-Danso et al. relied on integrating lentiviral vectors to express miRNA clusters for the efficient derivation of iPSC, Miyoshi et al. were able to deliver synthetic miRNA to successfully reprogram somatic cells, albeit at a low efficiency (Miyoshi, Ishii et al. 2011, Anokye-Danso, Trivedi et al. 2012).

In a very fruitful collaboration with Stemgent and by combining miRNA and mRNA reprogramming approaches, I established a non-integrative reprogramming system that is faster, easier and applicable to reprogramming a range of patient samples and for which I was selected for the spotlight program of this technology's leading provider (https://www.stemgent.com/campaigns/interview_with_sina_atashpaz). The inclusion of microRNAs cocktail (microRNA Booster Kit, Stemgent) in addition to the 5 factor mRNAs cocktail, accelerates the process to less than two weeks and supports the reprogramming of patient fibroblasts that are refractory to other methods. The removal of a feeder layer from the reprogramming process improves the visibility of morphological changes occurring in the target cells. The main advantage of the miRNA enhanced reprogramming method over the mRNA based method is shown in Fig. 1.8. The efficacy of the different protocols for generation of the iPSC is also shown in Fig. 1.8.



B

Conventional virus based method	miRNA/mRNA based approach
Low efficiency (0.01-0.1)	High efficiency (0.5)
Takes almost 30 days	Rapid reprogramming (10 days)
Integration of virus cassette	Eliminates the risk of genomic integration
Viral mediated insertional mutagenesis	Virus free
Hampering differentiation potency of iPSC	No need for virus excision or silencing
Need inactivated MEF layer	Eliminate adaptation steps
Refractory samples	Reprogramming recalcitrant samples

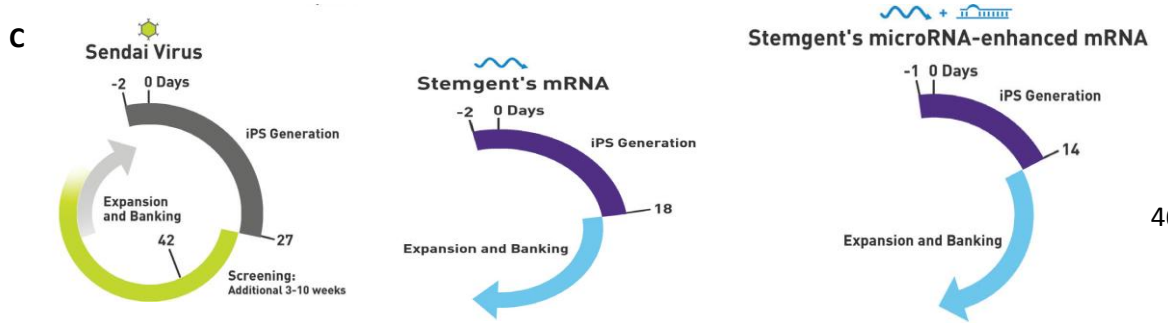


Fig. 1.9 A) Comparison of standard reprogramming methodologies.

mRNA proves to be the most efficient and safe reprogramming method as compared to all viral/DNA-based methods B) Comparison of conventional virus based reprogramming methods with miRNA enhanced reprogramming approach. C) RNA reprogramming experimental timelines compared to other reprogramming systems.

(Adapted from Stemgent mRNA Reprogramming System manual, http://assets.stemgent.com/files/1357/original/RNAReprogrammingProductSheet_final-full.pdf).

2.5.4.9.Variations in iPSCs

Genomic and epigenomic variabilities across reprogrammed iPSC lines have been recently recognized as a key concern for iPSC-based cell replacement therapy, disease modeling and drug discovery. In terms of disease modeling and drug discovery this could invite caution in the interpretation of results from few lines that do not adequately sample variability either across individuals or across lines reprogrammed from the same individual. The sources of such variability could be investigated at three levels: I) variability in the criteria that define iPSC lines as pluripotent cells, II) variability in cell lines from different donors (cell line variability due to different genetic backgrounds), and III) variability in cell lines from the same donor (clone variability due to the stochastic nature of the reprogramming method) (Vitale, Matigian et al. 2012, Liang and Zhang 2013). Many studies have recently tried to investigate the source of variability within each level. However, the precise contribution of each of the involved element has not been studied in a comprehensive study. This could be in part due to complexity and unknown mechanism of reprogramming process. More importantly such variations could be the results of the various parameters such as genetic and epigenetic alterations which can be involved at different level of the programming process. Among the genetic variations, aneuploidy, subchromosomal copy number variation

(CNV), and single nucleotide variations (SNVs) are the main ones while variations in X chromosome inactivation, variations in local epigenetic status (e.g., histone modification and DNA methylation) are considered as the key epigenetic variation that exists in PSC and as mentioned earlier, these variations could originate from different stages during the reprogramming process such as starting cells, during reprogramming and finally during passaging and culturing iPSC cells (Fig. 1.9) (Cahan and Daley 2013, Liang and Zhang 2013).

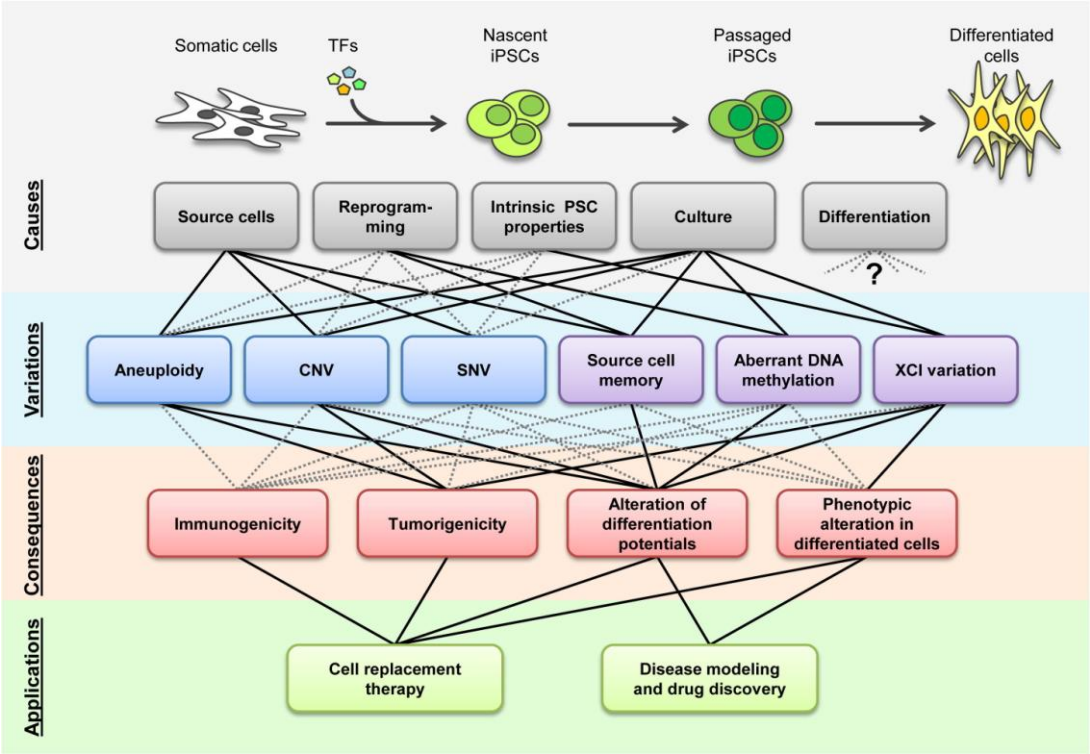


Fig. 1.10 Genetic and epigenetic variations and their causes, functional consequences, and impacts on applications iPSCs

derived from transcription factor (TF)-mediated reprogramming may bear different types of genetic (blue boxes) or epigenetic (purple boxes) variations that can be introduced from varied sources (gray boxes) during the derivation and manipulations of iPSCs. These variations may lead to different functional consequences (red boxes) that need to be considered when iPSCs or their derivatives are used for applications (green boxes). Solid lines, reported or definite connections; dotted lines, potential connections (Adapted from (Liang and Zhang 2013))

So far, two main methods have been suggested to reduce these variations. One is to minimize the epigenetic and genetic variation and the second one is the extensive characterization of each given iPSC line in order to exclude the lines which do not fell into the iPSC categories. Regarding the first criteria, reprogramming cells with the minimum genetic accumulation should be taken into the account. Moreover, non-integrating methods for the introduction of pluripotancy factors which can protect the genomic integrity and maintain epigenetic fidelity during reprogramming could be applied and finally, inclusion of some chemical that has also been shown to hamper the epigenetic aberrancy in iPSC shown to be useful. In terms of the second solution that is the careful monitoring and characterization of the variations a wide range of the analysis can be performed ranging from the basic characterizations such as karyotyping and expression analysis of the pluripotancy markers to the functional analysis including embryoid body and teratoma formation. In addition, more sophisticated and comprehensive analysis like genome-wide sequencing, expression analysis, and DNA and histone modification analysis could be exploited for genetic and epigenetic profiling of iPSC lines (Cahan and Daley 2013, Liang and Zhang 2013).

2.5.4.10. Differentiating hiPSCs to disease affected cell types

So far, most of the disease phenotype has been observed only in lineage differentiated cell types and not in iPSC. Therefore, gaining information on the pathogenesis of diseases could be mostly possible through the reliable differentiation protocols which enable the differentiation of the iPSC toward disease relevant cell types (Saha and Jaenisch 2009). During the past years differentiation of the iPSC into various cell types has already been achieved: neural progenitors (Chambers, Fasano et al. 2009), cortical neurons (Shi, Kirwan et

al. 2012, Shi, Kirwan et al. 2012), hepatocytes (Sullivan, Hay et al. 2010), blood cells (Choi, Yu et al. 2009), and neural crest(Lee, Chambers et al. 2010, Menendez, Kulik et al. 2013).

In order to better recapitulate the differentiation events in the developing embryo, most of these protocols take advantage of the small molecules and morphogens that were shown to be involved in development in vivo and can have either agonistic or antagonistic role in signaling pathways. Such molecules need to be used in specific concentration, sequence and correct time frame to induce the stepwise progression through a developmental program (Fig. 1.10). To achieve this goal, some protocols has exploited an alternative method in which they co-culture the target cells with the other cell types producing important factor to facilitate the differentiation process (Menendez, Kulik et al. 2013).

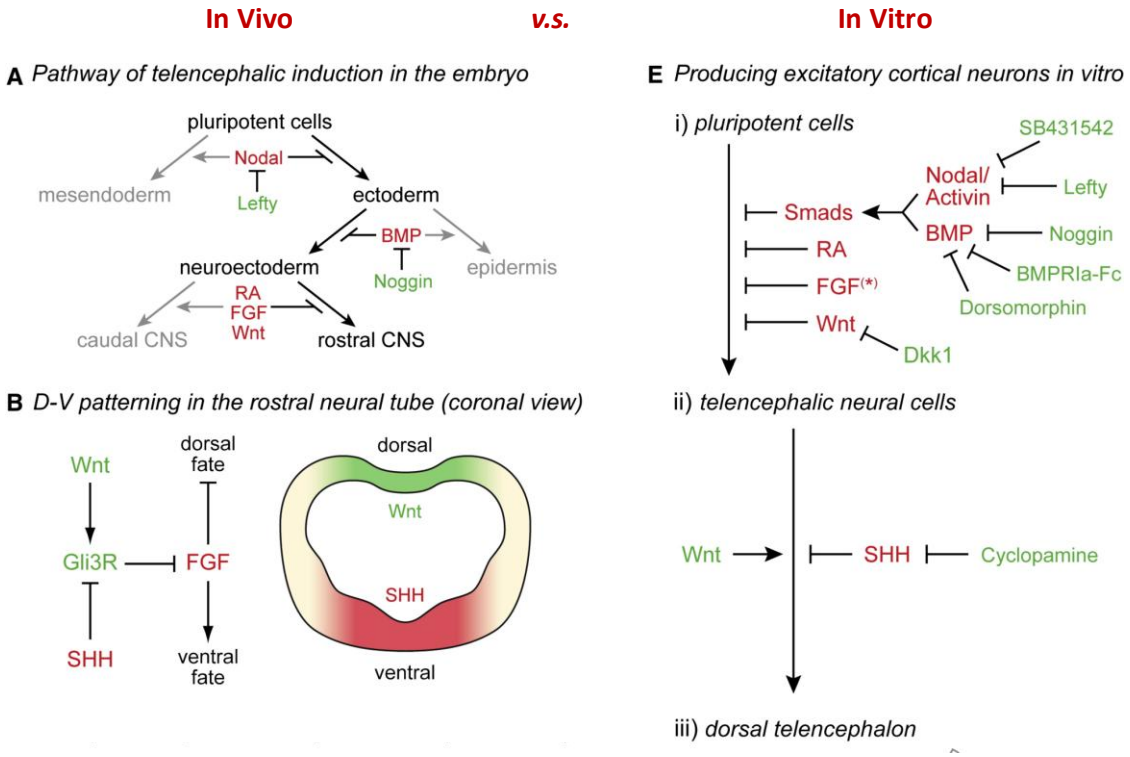


Fig. 1.11 Pathways for generating cortical excitatory neurons from pluripotent cells in vivo and in vitro.

Pluripotent cells of the inner cell mass in the embryonic blastocyst are thought to differentiate into cells of the anterior neuroectoderm in the absence of any instructive signals through a series of default fate decisions. Shown in red are morphogens that promote alternative differentiation fates. Shown in green are factors that inhibit those morphogens and facilitate the default pathway. (Modified from (Hansen, Rubenstein et al. 2011)).

Given the importance of the cerebral cortex in diseases such as epilepsy, autism and alzheimer's disease, the generation of the cortical neurons from iPSC is of great interest. However, the establishments of a protocol which can give rise to the various cell types within cortex need a deep understanding of this system. The cerebral cortex is composed of two main classes of neurons: the majority (80%) is excitatory glutamatergic projection neurons that are generated by various pools of cortical stem and progenitor cells. In brief, during early stages of neocortical development, neuroepithelial cells divide symmetrically to expand the progenitor pool before differentiating to radial glial cells (RGCs). RGCs however, divide asymmetrically not only to self-renew but also to produce either neurons or intermediate progenitor cells (IPCs). IPCs undergo symmetrical cell division to generate neurons, or additional IPCs. Basal RGCs (bRGCs) are similar to RGCs as they have a basal attachment at the pial basement membrane, but in contrast , do not maintain an apical process and thus have their cell bodies located in the outer margins of the SVZ. bRGCs also self-renew and generate IPCs and neurons. Interestingly, at the end of neurogenesis, RGCs and bRGCs transform into astrocyte progenitors (Fig. 1.12) (Franco and Muller 2013). Finally, cortical stem cells are believed to generate the six layers of the adult cortex in a stereotyped temporal order, with deep layer neurons being produced first and upper layer neurons being produced last. Of note, cortical neurogenesis lasts over 70 days in human while it takes only 6 days in mice (Shi, Kirwan et al. 2012, Franco and Muller 2013).

The second class of the cortical neurons are GABAergic interneurons (about 20%) that are generated outside of the cortex and migrate in during development. One of the major problems in generating cortical neurons is the reliable reproduction of the complex stem and progenitor cell populations found in the cortex in vivo. This is specially true because as discussed above, although neuroepithelial ventricular zone cells are the primary stem and progenitor population of the cerebral cortex, there are at least two secondary progenitor cell populations which have been identified in mouse, ferret and humans namely; basal progenitor cells and outer radial glial cells (oRG) (Fig. 1.11 A). All these groups of stem and progenitor cells appear to generate projection neurons. In addition the contribution of the oRG cells to increased size of the human cortex and diversification of the upper layer neurons has been proposed in recent studies (Franco and Muller 2013). On the basis of our current understanding of cerebral cortex development the Livesey group took advantage of two small molecules (i.e., SB431542 and Noggin) to inhibit Smad signaling pathway and developed a robust and efficient protocol that recapitulates crucial stages in human cortical development from PSC. This protocol consists of three main steps i) the directed differentiation of human PSCs to diverse population of cortical stem and progenitor cells, ii) an extended period of cortical neurogenesis, iii) a late phase of neuronal terminal differentiation to acquire mature electrophysiological properties, synaptogenesis and network formation (Fig 1.11B). More importantly, unlike differentiation of mouse ES cells, the diversity of cortical projection neurons in this system is roughly equal in terms of the deep (early born) and upper (late born) layer neurons (Shi, Kirwan et al. 2012, Shi, Kirwan et al. 2012).

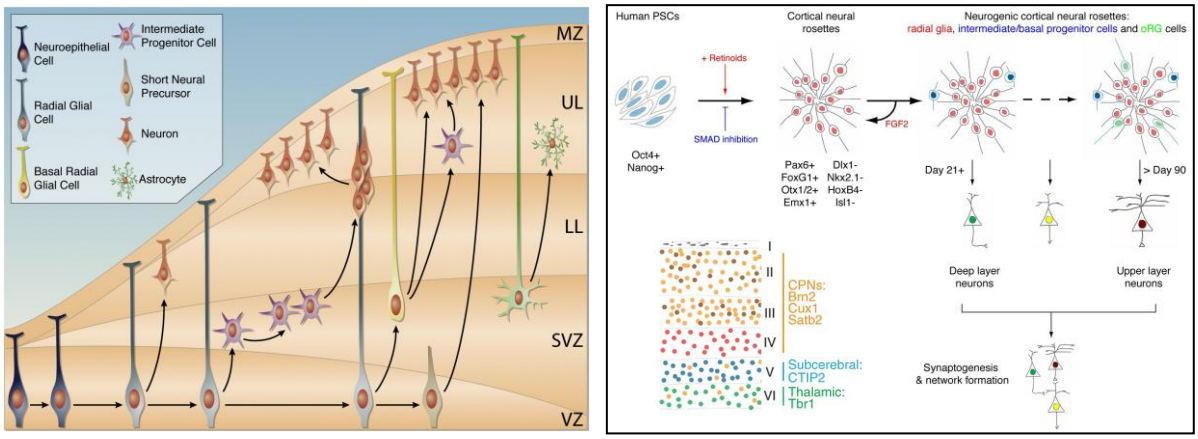


Fig. 1.12 Subtypes of stem and progenitor cells in the developing neocortex.

A) A Schem Showes the diversity of the stem and progenitor cell during cortical embryonic development which undergo severl symetrical and asymetrical cell division to generate wide range of lower and upper layer neurons. Adapted from (Franco and Muller 2013)

B) Schematic of the cortical differentiation protocol reported by the Levisey group through the combination of dual SMAD inhibition, combined with retinoids which is able to differentiate PSCs to cortical stem and progenitor cells that can be expanded/maintained with FGF2. The removal of FGF2 differentiate Pax6-expressing radial glia, Tbr2-expressing basal progenitor cells and Pax6+/Tbr2- outer radial glia (oRG) cells into lower layer neurons generated early, beginning around day 21 and upper layer neurons generated last, continuing beyond day 90. (Modified from (Shi, Kirwan et al. 2012)).

In a different attempt, Menendez et al. developed a highly efficient, lineage-specific differentiation of human pluripotent cells to NCSC fate which can be applied to the modeling of neural crest–related human diseases (Menendez, Kulik et al. 2013). Neural crest stem cells are a transient population arising from the neural plate border which upon delamination from the roof plate can migrate to different regions of the embryo and accordingly differentiate into a wide range of cell types such as sensory neurons, Schwann cells, melanocytes, cells that make up the craniofacial structures such as bone and cartilage and finally smooth muscle cells

that contribute to the heart valves (Lee, Chambers et al. 2010, Menendez, Kulik et al. 2013) (Fig. 1.12) . To this aim, the authors combined Smad inhibition with activation of the Wnt pathway to develop a single-step, highly efficient method for the generation of NCSCs from hESCs and hiPSCs. The key advantages of this method are that first, unlike previously described methods, this protocol does not require a co-culture on feeder layers to generate neural crest. Second, the efficiency of this protocol is very high and therefore does not require FACS sorting to obtain enriched populations and finally the generated NCSC using this method has been shown not only to have self-renewal potential upon freezing, thawing and multiple passaging but more importantly they can be differentiated into a wide range of differentiated cell types including smooth muscle cells, adipocytes, chondrocytes, osteocytes and peripheral neurons (Menendez, Kulik et al. 2013).

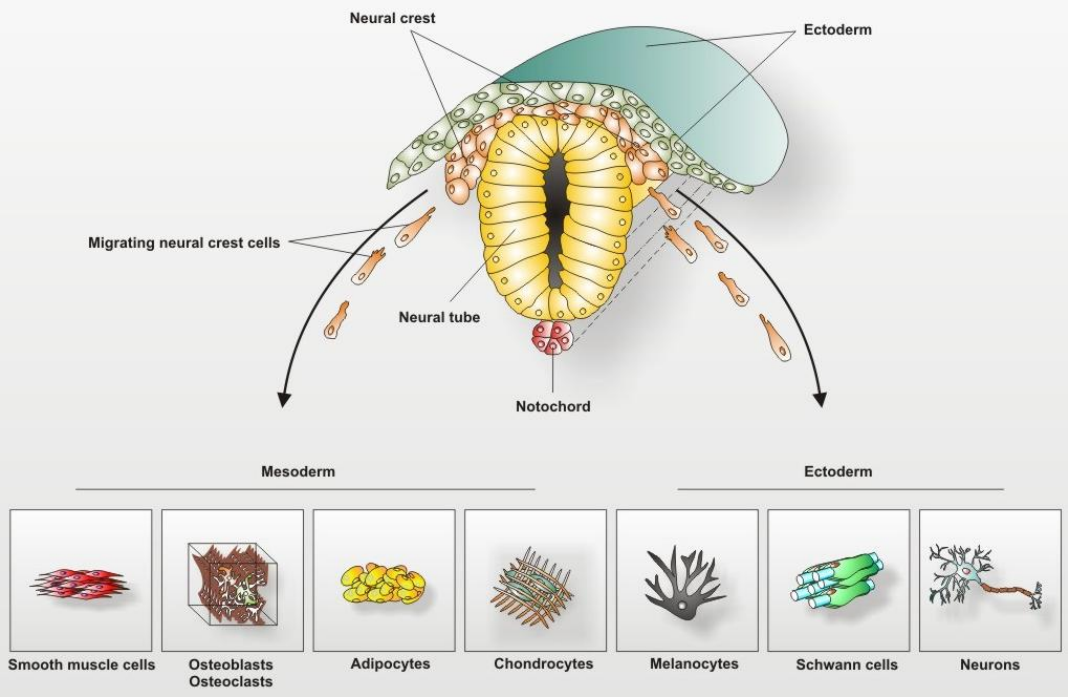


Fig. 1.13 Embryonic development of neural crest stem cells.

After neurulation neural crest cells migrate out and differentiate into multiple cell types like peripheral neurons, adipocytes, smooth muscle cells and Schwann cells. adapted from

http://web.biologie.uni-bielefeld.de/cellbiology/index.php/research/neural-crest-derived-stem-cells

Taking advantages of two key establishments that I have implemented in the lab namely mRNA based reprogramming and neural differentiation protocol, in my PhD project, I aimed to understand two key questions in the context of modeling 7 q11.23 copy number variation among large cohort of samples: i) the extent to which early developmental lineages are informative about disease-relevant pathways affected by genetic mutations and, ii) the feasibility of reliably identifying those pathways beyond the sources of variability inherent to the iPSC-based approach.

3. Materials and Methods

3.1. Human samples

Participation in this study by patients and relatives along with skin biopsy donations and informed consent procedures were approved by the Ethics Committee of the Genomic and Genetic Disorder Biobank (Casa Sollievo della Sofferenza, San Giovanni Rotondo, Italy) and the University of Perugia (Azienda Ospedaliera-Universitaria "S. Maria della Misericordia", Perugia, Italy).

3.2. Fibroblast culture and reprogramming

Primary fibroblast cell lines WBS1-2-3-4, 7Dup-ASD2, AtWBS1, CTL1R were obtained from Genetic Disease Biobank. 7Dup-ASD1 primary fibroblast was obtained from Azienda Ospedaliera-Universitaria "S. Maria della Misericordia", Perugia, Italy. Fibroblasts were cultured in HF medium composed as follows: RPMI 1640, 1% L-Glutamine, 1% Pen-Strep, 15% FBS for few passages before reprogramming.

WBS1-2-3-4, 7Dup-ASD1-2, AtWBS1, CTL1R, and CTL2 fibroblast lines were reprogrammed using mRNA Reprogramming Kit (Stemgent) with some modifications. Briefly, 4-5x106 million newborn foreskin fibroblast cells (NuFF) (Stemgent) were plated onto a 75-Tflask with Plurition media (Stemgent) over 8 days and media collected daily and used as condition Pluriton media during reprogramming. Next, target fibroblast cells were plated at different cell densities (5000-10,000) onto already Nuff plated plates. Cells pre-incubated with B18R (Stemgent) and transfection with the mRNA cocktail (OCT4, KLF4, CMYC, LIN28 and SOX2) along with nuclear GFP (Stemgent) were perform daily for about 16 days. For reprogramming 7Dup-ASD2 and CTL1R lines, microRNA Booster Kit

(Stemgent) were used to enhance the reprogramming. To this goal, I followed the mRNA reprogramming protocol with the following modifications: i) Target cells were plated onto matrigel coated plates instead of NuFF cells. ii) Target cells were plated at higher cells densities i.e., 5-10x104. iii) Cells were transfected with miRNA cocktail (Stemgent) at day 1 and 5 of reprogramming. CTL3 and WBS2 line was reprogrammed using STEMCCA polycistronic lentiviral vector CTL3 underwent cre-mediated excision of the integrated polycistron. For mRNA-mediated reprogramming epithelial-to-mesenchymal transition was monitored from day 5 by tracing GFP-positive cells. Successfully reprogrammed colonies were assayed for pluripotency at day 20 using a live TRA-1-60 antibody (Stemgent) and selected for further expansion as detailed below. All cells culture at low O2 tension (5%).

3.3.iPSC culture

iPSC lines were cultured on mitomycin C-inactivated mouse embryonic fibroblasts (MEFs) as previously described (Takahashi, Tanabe et al. 2007) in a medium composed as follows: DMEM-F12 (Gibco) in a 1:1 ratio supplemented with 20% KSR, 1% Non Essential Amino Acids, 1% Pen-Strep, 1% Glutamine, 0.1 % beta-mercaptoethanol, 10 ng/ml basic fibroblast growth factor (bFGF, Gibco). Colonies were passed and expanded twice by physical fractionation with a sterile needle and re-plated onto newly seeded MEFs for line establishment. After few passages iPSCs were adapted to grow in feeder-free condition on plates coated with human-qualified Matrigel (BD Biosciences) diluted 1:40 in DMEM-F12 and in mTeSR-1 (StemCell Technologies) medium and were passed by physical fractionation upon a 2 minutes treatment with Dispase (Sigma) at 37°C. Feeder-free iPSCs were also adapted to grow in single cell culture by dissociating them by a 3 minutes treatment with

Accutase (Sigma) at 37°C and finally resuspended in a suitable volume of mTeSR-1 supplemented with 5 μM Y-27632 (Sigma).

3.4.Teratoma assay and immunohistochemistry

Teratoma assay was performed by subcutaneously injecting 1-3x106 iPSCs in human-qualified Matrigel (BD Biosciences) into the dorsal flanks of NOD-SCID IL2RG male mice. Teratomas were isolated when the diameter reached > 1.5 cm and fixed in 4% buffered formalin. Samples were then OCT embedded, sectioned and stained for H&E and germ layer specific antibodies: desmin (Dako), S-100 (Dako) and cytokeratin (Dako).

3.5. Immunocytochemistry

Cells were fixed in 4% PFA for 20' and subsequently blocked in 10% FBS + 0.1% Triton for 30' at room temperature. Cells were incubated with primary antibodies overnight at 4°C and then with secondary antibodies for one hour at room temperature. Primary and secondary antibodies were resuspended in 10% FBS. Primary antibodies used were OCT3/4 (SantaCruz), NANOG (Everest Biotech), SSEA3 (Invitrogen), Tra1-60 (Stemgent), TBR2 (Abcam), SOX2 (R&D), PAX6 (HBDS), NESTIN (Abcam), FOXG1 (StemCulture), ZO1 (Invitrogen), OTX2 (Millipore), Ki67 (Abcam), PHH3 (Millipore), CLIP2 (Abcam), BRN2 (Santa Cruz), TUJ1 (Covance), vGlut1 (Synaptic systems), MAP2 (Covance), vGLUT2 (Synaptic System), GAD67 (Millipore), Cux 1(Santa Cruz), Satb2 (Abcam), TBR1 (Abcam), GFAP (DAKO). Alkaline phosphatase staining was performed using Alkaline Phosphatase Detection Kit (Sigma). Images were acquired at an Olympus AX70 microscope.

3.6.DNA, RNA and protein extraction

Genomic DNA was extracted from fibroblasts and feeder-free iPSC lines using the DNeasy Blood and Tissue Kit (Qiagen) according to manufacturer specifications. RNA was extracted from iPSC lines using the RNeasy Micro Plus Kit (Qiagen) according to manufacturer specifications, substituting the genomic DNA elimination column by needle and Dnase treatment (Qiagen). Quality and concentration of DNA and RNA was assessed using a NanoDropSpectrophotometer (NanoDrop Technologies).

Proteins were extracted as follows: cells were scraped from the plate and centrifuged at 1100 g at 4°C for 3 minutes, then washed in PBS and lysed in RIPA buffer plus protease inhibitors cocktail (Sigma) on a spinning wheel at 4°C for 30 minutes. Lysates were sonicated using the Bioruptor Sonication System (UCD200) for 3 cycles of 30 seconds with 60 seconds breaks at high power. Lysates were centrifuged at 13000 g for 15 minutes and supernatants were transferred to a new tube. Protein quantification was performed using the Bradford protein assay (BioRad) and following manufacturer instructions.

3.7. Immunoblotting

For immunoblotting 20 to 40 μg of protein extract per sample were run on a precast Nupage 4-12% Bis-tris Gel (Life Technologies), transferred on a nitrocellulose membrane and blocked in TBS-T and 5% milk. Antibodies used for detection were GTF2I (Cell Signaling), BAZ1B (Abcam) and GAPDH (Abcam), OCT3/4 (SantaCruz). Blots were scanned using a LI-COR Odyssey Infrared Imaging System and bands were quantified using ImageJ software.

3.8. RNAseq and Nanostring

Library preparation for RNA sequencing was performed using Poly-A, RiboZero and Single Stranded kits (Illumina) according to manufacturer instructions.

Nanostring quantification was performed according to manufacturer instructions and data normalization was performed using the nSolver Analysis Software 1.1.

3.9. RNAseq analysis

Reads were aligned to the hg19 transcriptome using TopHat 2.0.10. The alignment was first performed on the RefSeq transcriptome and all the reads that had an edit distance ≥ 1 were realigned on the genome, allowing a maximum read edit distance of 3 and 3 (100 bp reads) and 2 (50 bp reads) maximum mismatches. 50bp stranded reads were analyzed using the "fr-firststrand" option. Quantification of reads over the RefSeq transcriptome was performed with Cufflinks 2.2.1 using sequence-bias and multi-read corrections. Differential gene expression was estimated using Cufflinks 2.2.1, using per-condition dispersion models.

For the iPSC stage, given the presence of both polyA and Ribo-zero samples, we considered the union of DEGs identified through a global analysis of all samples (FDR < 0.05) with those identified through independent analysis of the polyA and Ribo-zero samples. In the latter case, we considered as differentially expressed genes that had a FDR < 0.2 in both datasets (comparing the same genotypes), and for which the change was in the same direction.

3.10. Down sampling test

The majority of differentially-expressed genes identified in this study were still found when removing the external controls, and most of the remaining DEGs were close to

significance, arguing against the introduction of a major bias through the use of external controls.

In order to assess the effect, on the transcriptional analysis, of having fewer samples, we repeated the analysis of our polyA dataset (focusing on the comparison for which we had the most samples, i.e. the global analysis of WBS vs CTL iPSC), using only subsets of the samples. Random removal of 1 clone per patient lead to a dramatic reduction in the number of DEGs (48 to 76% lost), and to the identification of DEGs that are falsified by the discarded data. The impact of removing all clones from one patient per condition (amounting to fewer samples than removing one clone per patient) was even greater. In contrast, depth of sequencing appeared to make little difference: reducing coverage by half led to the loss of 11% of DEGs and to very few false positives.

3.11. Shuffling tests

To assess the possibility that the observed differential expression might arise due to random variations, we performed a series of differential expression analysis between randomly-selected samples, discarding comparisons in which the two groups were not balanced for sex and/or genotype. A minimum of 3 such combinations were tested per tissue, and the resulting genes were pooled for the purpose of enrichment analysis.

At the iPSC stage (using the polyA dataset), we first randomly assigned all patients to two groups, but we could obtain statistically significant genes in none of the combinations. We therefore gradually removed patients until significant genes were obtained, which did not happen until a comparison involving 6 vs 6 samples (in each group, 3 samples from 2 patients). In contrast, when clones were selected and assigned to groups in a way that maximized the number of patients represented in each group, we had to go down to 3 vs 3 (3

samples per group, coming from 3 different patients) to get statistically significant genes (18 DEGs, showing no significant GO enrichment). Similarly, random allocation of the control samples (including external controls, balanced across groups) yielded very few DEGs (maximum 17) and no significant GO enrichment.

These results suggest that the primary source of "spurious" differential expression is genetic variation between individuals, which only gets mitigated using lines derived from several patients.

Finally, it is interesting to note that despite yielding very few DEGs (maximum 78), some of the 6 vs 6 comparisons showed statistically significant enrichment for the GO categories of extracellular matrix organization and extracellular structure organization, pointing to these genes as particularly varying in expression between lines and/or individuals.

In differentiated cell types, shuffling tests using (3 combinations of) 3 vs 3 samples yielded either no significant differential expression (NCSC), or very few genes (NPC) that displayed no significant enrichments, with the exception of the MSC dataset. The union of genes found significant from random combinations of the MSC samples showed several GO enrichments, including (albeit at a lower level) some categories that were found significant between the genetic conditions. However, removal of these genes did not significantly alter the main categories enriched among the DEGs between genetic conditions.

3.12. cDNA preparation and qPCR

Retrotranscribed cDNAs have been obtained from 1 µg of total DNA-depleted RNA using the superscript VILO retrotranscription kit from Life Technologies according to manufacturer instructions.

For real time q-PCR analysis a total amount of cDNA corresponding to 5 ng of starting RNA has been used for each reaction. FAST SYBR green master mix from Life Technologies and 10 μM primers pair have been used. The qPCR reactions have been performed on an Applied Biosystems® 7500 Real-Time PCR machine following the standard amplification protocol.

The pair oligos used for qPCR were: GAPDH (F: GCACCGTCAAGGCTGAGAAC, R: AGGGATCTCGCTCCTGGAA), SOX10 (F: CTTCATGGTGTGGGCTCAG, R: GCTTGTCACTTTCGTTCAGC) and GTF2I (F: GATCTTGCAACCCTGAAATGG, R: CACCTGGAGATAGTATTGACCTG). SOX9 (F: CACAGCTCACTCGACCTTG, R: ACACAAATGTCCAAAGGGAATTC). ZIC1 (F: CCTACACGCATCCCAGTTC, R: TTGTGGTCGGGTTGTCTG). TFAP2 (F: GTTACCCTGCTCACATCACTAG, R: TCTTGTCACTTGCTCATTGGG). NGFR (F: GTGGAGAGTCTGTGCAGTG, R: ATCGGTTGTCGGAATGTGG). ELN (F: CCTGGCTTCGGATTGTCTC, R: CAAAGGGTTTACATTCTCCACC).

3.13. Lentivirus production

Production of STEMCCA and FOXG1 lentiviral particles was performed as previously described (Sommer, Stadtfeld et al. 2009). Briefly, plasmids expressing viral proteins GAG, POL, REV, TAT, and the vesicular stomatitis virus envelope glycoprotein (VSV-G) were co-transfected with STEMCCA vectors into semi-confluent 293T cells by calcium phosphate precipitation. For the Syp-GFP production second generation packaging system (PAX and VSV-G) were used. For Supernatant of transfected cells were collected every 12 hours during 2 consecutive days and concentrated by centrifugation. Viral particles were resuspended in iPSC medium and either used freshly for infection or frozen at −80°C.

3.14. Differentiation

Differentiation of iPSC cells toward cortical neurons were carried out based on the previously published protocol (Brennand, Simone et al. 2011). Differentiation into the dorsal telencephalic lineage was accomplished by dual Smad inhibition in the presence of SB431542 (Tocris) and Noggin (R&D) (Shi, Kirwan et al. 2012, Shi, Kirwan et al. 2012) with some modifications. Briefly, iPSC cells were adapted to single cells culture and plated at the optimal cells density onto matrigel coated plates to have 90% confluency the next day. Cells were treated with the inhibitors for 10-12 days and then re-plated using Dispase to initiate the formation of rosette like structures. Next, cells were splitted using Accutase to re-plated onto polyornitin-laminin coated plated at the right cell density for terminal differentiation. iPSC Differentiation of NCSC and MSC was performed as previously described (Menendez, Kulik et al. 2013), through the activation of Wnt signalling and Smad pathway blockade by administering the small molecules GSK3i (Calbiochem) and SB431542 (Tocris).

3.15. Flow cytometry

$1x10^6$ cells were fixed in 4% PFA and subsequently blocked in 10% BSA. Cells were incubated for one hour with primary conjugated antibody resuspended in 1-2% BSA. The primary conjugated antibodies used were CD57-FITC (HNK1, BD), CD271-647 (NGFR, BD), CD44-APC (EBIOS) and CD73-PE (BD). Analysis was performed on FACSCalibur (BD Biosciences) and data were analyzed with FCS express software (Tree Star Inc.).

3.16. CGH array

DNA was isolated from parental fibroblast and iPSC using Qiagen kit as described above. DNA concentration and purity were determinate with a ND-1000 spectrophotometer (NanoDrop Technologies, Berlin, Germany) while whole-genome copy number variations

(CNVs) analysis was carried out using the CytoScan HD array platform (Affymetrix, Santa Clara, CA). The CytoScan HD assay was performed according to the manufacturer protocol, starting with 250 ng of DNA. Briefly, total genomic DNA was digested with a restriction enzyme (NspI), ligated to an appropriate adapter for the enzyme and subjected to PCR amplification using a single primer. After digestion with DNase I, the PCR products were labeled with a biotinylated nucleotide analogue, using terminal deoxynucleotidyl transferase (TdT) and hybridized to the microarray. Hybridization was carried out in the Hybridization Oven 645 while subsequent washing and staining were performed using the Fluidics Station 450.

3.17. CGH array analysis

Each array was then scanned with the Scanner 3000 7G and both quality control step and copy number analysis were performed using the Chromosome Analysis Suite Software version 2.0: i) the raw data file (.CEL) was normalized using the default options; ii) an unpaired analysis was performed using as baseline 270 HapMap samples in order to obtain Copy numbers value from .CEL files while the amplified and/or deleted regions were detected using a standard Hidden Markov Model (HMM) method.

3.18. Microarray

Microarray analysis was performed with the Affy package using Marc Carlson's Hugene 2.1st RefSeq annotation file, version 18. Background normalization was performed using the RMA method, whereas between-sample normalization was performed using the quantile normalization method. Quantification of expression was obtained using perfectly matching probes only with median polish summarization. Probe sets not assigned to known genes or having log2 fold changes < 0.5 were discarded and differential expression was

assessed using a 2-tailed t-test. For the purpose of enrichment analyses, DEGs with a FDR < 0.2 were considered.

3.19. Gene ontology enrichment analysis

For RNAseq data, enrichment analysis was performed using the R package GO seq in order to correct for transcript length bias considering only categories with at least 10 annotated genes and discarding categories that had less than 8 significant genes. For genes measured by other methods, the enrichment analysis was performed with the package TopGO using the classic algorithm and Fisher's test with the same cutoffs described above. In order to create enrichment treemaps, parent categories that had enriched children were first removed and then maps were created with the package Treemap, using as colors the combination of non-overlapping parent categories accounting for the largest proportion of plotted categories. All reported FDR values were calculated using the Benjamini-Hochberg method.

4. Results

4.1. Recruitment of a large cohort of WBS and 7dupASD patients

In collaboration with the Genomic and Genetic Disease Biobank (http://www.telethon.it/en/scientists/biobanks) supported by the Telethon foundation, we selected a unique combination of skin fibroblast from a large cohort of patients affected by WBS and 7dup-ASD along with the fibroblasts from their healthy relative as half matched controls (Table1) in one case as in the absence of disease-specific iPSC lines with isogenic controls, such control iPSC lines from unaffected parents are essential for identifying relevant phenotypes. Patients were assessed by a multidisciplinary team of specialists for a detailed record of clinical, functional, behavioral information (Tab. 3.1). This cohort includes: i) five patients carrying the typical WBS deletion; ii) one patient carrying an atypical WBS deletion that spares several genes including BAZ1B, who exhibits milder craniofacial dysmorphisms and lack cardiovascular abnormalities, supporting a role for BAZ1B in neural crest-derived lineages; iii) two patients carrying the typical duplication of the 7q11.23 interval associated to language impairment, autism spectrum disorder and craniofacial dysmorphisms; and iv) two unaffected relative of a typical WBS patient, chosen as genetically half-matched control. All patients were already diagnosed at the molecular level by a combination of FISH, MLPA, and qPCR (as also shown in Fig.3.4B for the 7q11.23 aCGH profiles of representative case of each genotype).

Clinical features		GDB192/WBS154 AtWBS1	GDB306/WBS276 WBS4	GDB316/WBS301 WBS2	GDB361/WBS309 WBS3	GDB339/WB WBS1	GDB CF 7dupASD1	GDB242/WBS202 7dupASD2
Intellectual disability		Moderate	Moderate	Moderate	Moderate	Moderate	Moderate	Moderate
Cardiovascular	Supravalvular aortic stenosis	-	+	-	-	+	NA	NA
	Peripheral pulmonary stenosis	-	-	-	+	-	NA	NA
	Valvular pulmonic stenosis	-	+	-	-	-	NA	NA
	Hypertension	-	-	NA	-	+	NA	NA
	Others	-	-	-	left ventricular hypertrophy; interatrial septal defect	-	NA	NA
Craniofacial	Wide mouth	+	-	+	-	+	NA	NA
	Prominent ear lobes	-	-	-	+	+	NA	+
	Dolicocephaly	-	+	+	-	-	NA	NA
	Broad forehead	-	+	+	-	+	-	NA
	High narrow forehead	-	-	-	-	-	+	NA
	Microcephaly	-	+	+	+	+	-	NA
	Macrocephaly	-	-	-	-	-	+	NA
	Bitemporal narrowing	+	+	+	+	+	NA	NA
	Periorbital fullness	+	+	+	+	+	NA	NA
	Epicanthal folds	-	-	+	+	+	NA	NA
	Long eyelashes	NA	NA	NA	NA	NA	+	NA
	Stellate irides	+	+	+	+	+	NA	NA
	Malar flattening	+	+	+	-	+	NA	NA
	High broad nose	-	-	-	-	-	+	NA
	Short upturned nose	-	+	+	+	+	-	NA
	Bulbous nasal tip	+	-	+	+	+	NA	NA
	Bridge nose flattened	-	-	+	+	+	NA	NA
	Long philtrum	+	+	+	+	+	-	-
	Short philtrum	-	-	-	-	-	+	+
	Full lips	+	+	+	-	+	NA	NA
	Full cheeks	+	+	+	-	+	NA	NA
	High arched palate	-	-	-	-	-	+	NA
	Dental abnormalities/ malocclusion	-	+	+	NA	+	+	NA
	Small or unusually shaped primary teet	-	+	NA	NA	+	NA	NA
	Hypodontia	-	NA	NA	NA	+	NA	NA
	Small mandible	NA	NA	NA	NA	NA	+	NA
	Retrognathia	NA	NA	NA	NA	NA	+	NA
Endocrine	Subclinical hypothyroidism	-	+	+	NA	+	NA	NA
	Precocious puberty	-	NA	NA	NA	-	NA	NA
	Hypercalcemia	-	+	-	-	+	NA	NA
	Glucose intolerance or diabetes mellitu	-	-	NA	NA	-	NA	NA

Clinical features		GDB192/WBS154 AtWBS1	GDB306/WBS276 WBS4	GDB316/WBS301 WBS2	GDB361/WBS309 WBS3	GDB339/WB WBS1	GDB CF 7dupASD1	GDB242/WBS202 7dupASD2
Gastrointestinal	Feeding difficulties	-	+	-	+	+	NA	NA
	Abnormal weight gain	-	+	-	-	+	NA	NA
	Celiac disease	-	-	-	NA	NA	NA	NA
	Constipation	+	+	-	-	-	NA	NA
	Gastroesophageal reflux	-	-	-	-	-	NA	NA
	Abdominal pain (unclear cause)	-	NA	-	NA	NA	NA	NA
	Rectal prolapse	-	NA	-	NA	NA	NA	NA
	Diverticular disease	-	-	-	NA	-	NA	NA
Genitourinary	Congenital anomalies	-	+ *	-	+	-	NA	NA
	Enuresis	-	-	-	+	+	NA	NA
	Nephrocalcinosis	-	-	-	-	-	NA	NA
Ocular	Strabismus	+	-	+	+	+	NA	NA
	Hypermetropia	-	+	-	NA	-	NA	NA
	Narrowing of lacrimal duct	-	-	-	+	-	NA	NA
Musculoskeletal	Kyphosis	+	-	-	-	-	NA	NA
	Lordosis	-	-	-	-	-	NA	NA
	Scoliosis	+	-	-	-	NA	NA	NA
	Joint laxity	-	+	-	+	+	NA	NA
	Radioulnar synostosis	-	-	+	-	-	NA	NA
	Umbilical hernia	-	-	-	-	-	NA	NA
Neurological and neuro	Hypotonia	-	+	+	+	+	+	+
	Hyperrelexia	-	-	-	NA	-	NA	NA
	Cerebellar findings	-	-	-	-	-	NA	NA
	Type 1 Chiari malformation	-	-	-	-	-	NA	NA
	Hoarse voice	-	+	+	+	+	NA	NA
	Hyperacusis	+	+	+	NA	+	NA	-
	Sensorineural hearing loss	+ (mild, right ear)	-	+	NA	-	NA	-
	Epilepsy	-	-	-	-	-	NA	+
	Sleep dysregulation	-	-	-	NA	NA	NA	-
	Speech impairment	-	-	-	-	-	+	+
	Social impairment	-	-	-	-	-	+	-
	Stereotypies	NA	NA	NA	NA	NA	+	-
MRI anomalies	Cortical thickening	NA	NA	NA	NA	NA	+	+
	Ventricular dilatation	NA	NA	NA	NA	NA	NA	+
	Simplified gyral pattern	NA	NA	NA	NA	NA	+	NA
	Increased intracranial volume	NA	NA	NA	NA	NA	NA	+
	Decreased amygdala/intracranial volum	NA	NA	NA	NA	NA	NA	+
	No activation of emotion-processing are	NA	NA	NA	NA	NA	NA	+
Other	Recurrent otitis media	-	-	-	-	-	NA	NA
	Short stature	-	+	-	+	-	NA	NA
	Inguinal hernias	-	-	-	-	-	NA	NA
Reference		Fusco et al. EJHG 2014					Torniero et EJHG 200	Prontera et al. JADD 2014

Tab. 3.1 Clinical features of patients recriute in this study. NA, data not available. "+" stands for present; "-" stands for not present. * bilateral cryptorchidism, penile hypospadia

4.2. Establishment of lentiviral-based somatic cell reprogramming

First I established human somatic cell reprogramming using excisable polycistronic lentiviral vector (STEMCCA) in order to generate patient derived iPS cells (Fig. 3.1 A). This technology enables the excision of the viral transgenes in the reprogrammed clones by administration of Cre recombinase as the viral cassette is flanked by loxP sites. This could be important as the residual presence of the virus transgenes could generate unexpected long-term instability and tumorigenicity due to permanent genetic integration (Pasi, Dereli-Oz et al. 2011). Next, I characterized six independent clones using different in-vitro and in-vivo pluripotency assays (Fig. 3.1). Fig. 3.1 shows data from two representative iPSC lines from sample WBS2, in which I confirmed typical iPSC morphology (Fig. 3.1 B) and the expression of key pluripotency markers like NANOG, SSEA-3 and TRA-1-80 and alkaline phosphatase (Fig.3.1 C). Lentiviral induced PSC could from teratoma containing all three germ layers when they were injected into the flank of immmunodeficient mouse (Fig 3.1 E) Karyotyping analysis also has shown a stable cell line with a normal female karyotype (Fig. 3.1 F).

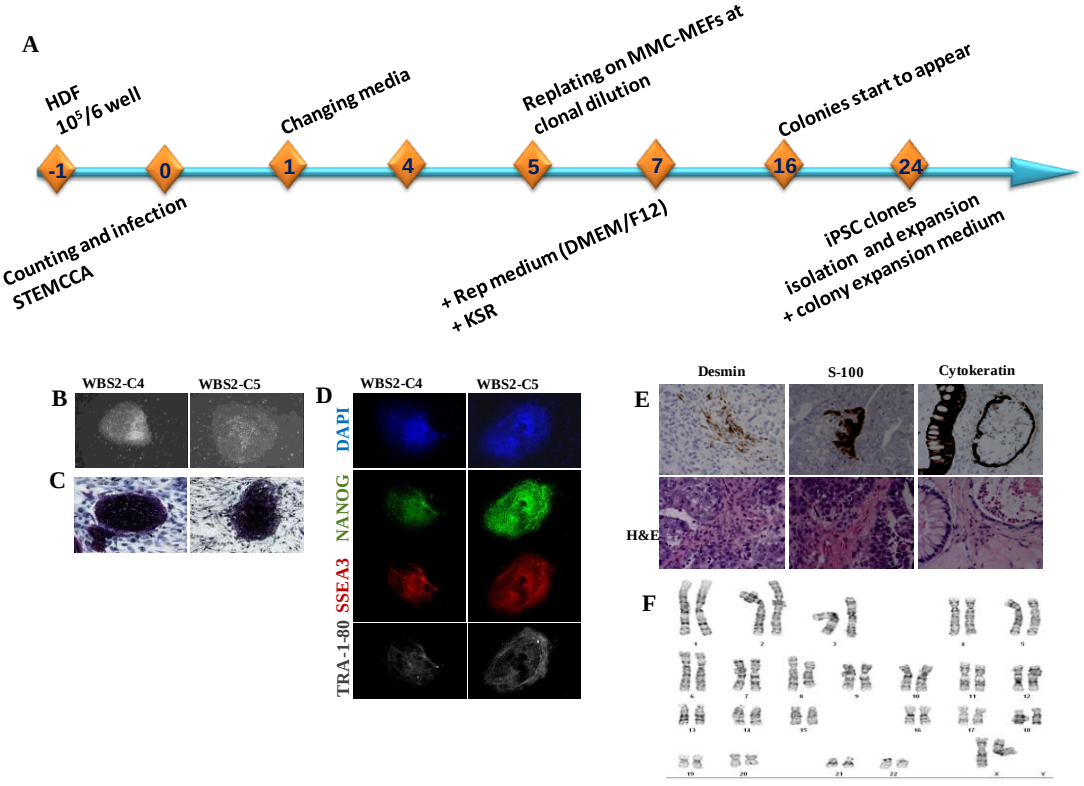


Fig. 3.1 Establishment and charactrization of lentiviral-based iPSC

A) Scheme showing the protocol used for the generation of iPSC using excisable polycistronic lentiviral vector. B) Morphology of established iPS lines by phase contrast microscopy. C) Expression of alkaline phosphatase by immunohistochemistry. D) Immunofluorescence detection of iPSC colonies expressing pluripotent markers NANOG (green), SSEA-3 (red) and Tra-1-80 (gray). Nuclei were stained with DAPI (blue). E) Teratomas containing all three germ layers were formed after subcutaneous injection of iPS cells into NOD-SCID mouse. F) Normal karotype of the patient derived iPSC clone.

4.3. Differentiation of the iPSC toward cortical neurons through embryoid body intermediates

I then subjected six iPSC lines into cortical neural differentiation, by modifying a previously published protocol that taking advantage of several steps including of embryoid body and neural tube-like rosettes structures that can be subsequently split in order to establish neural progenitor cell (NPC) cultures (Brennand, Simone et al. 2011). The modifications are detailed in Fig. 3.2 A. I obtained neural tube-like rosettes structures and NPC that express defining neural stem cell markers such as Nestin, Pax6 and Sox2 (Fig. 3.2B and C). Prior to proceed with the differentiation of NPC toward mature neurons; I tested whether the reprogramming transgene had been silenced, since sustained expression would have likely hampered terminal neuronal differentiation. To do this, I performed immunobloting for the Oct-4 in iPSC-derived NPC and iPSC and the results confirmed the efficient silencing of virus transgene upon neural differentiation (Fig. 3.2D). Importantly, this result showed that virus transgene excision is not required to ensure robust differentiation. Finally, I differentiated neural progenitor cell toward mature neurons for 70 days (Fig. 3.2E).

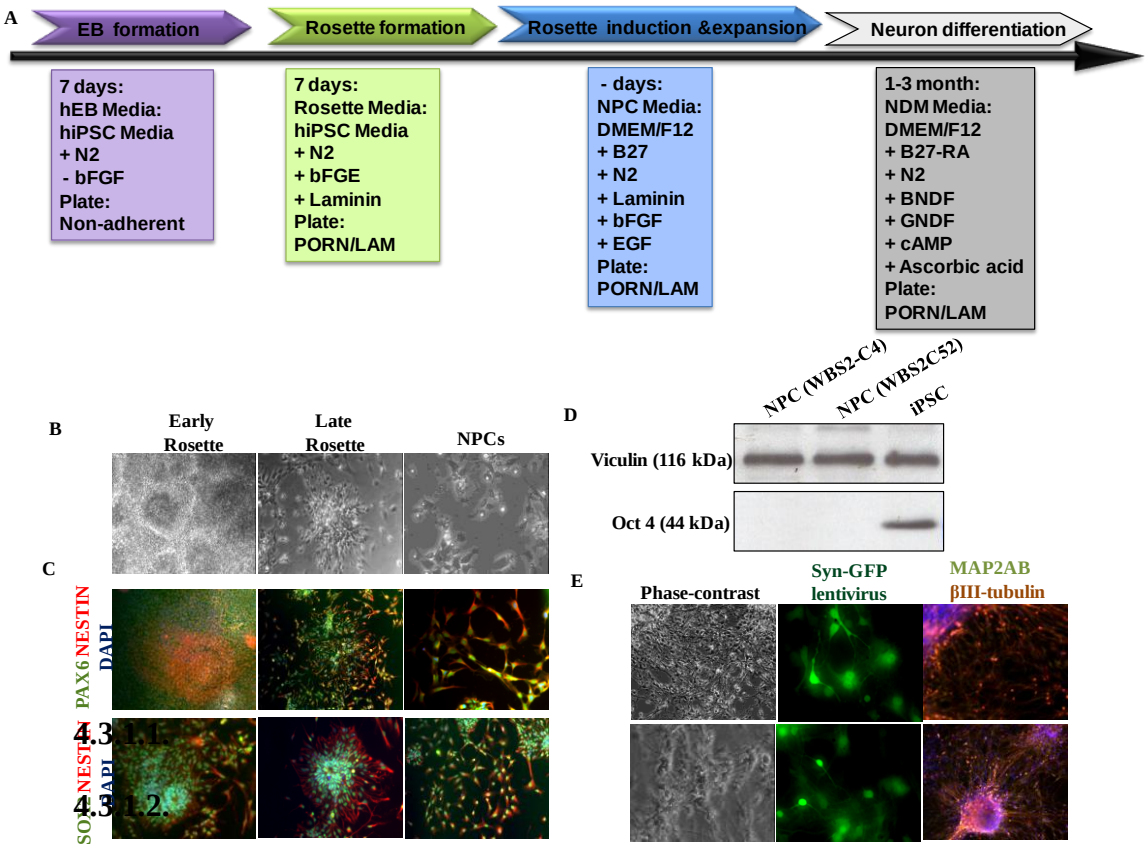


Fig. 3.2 Differentiation of the virus iPSC toward cortical neurons through embryoid body intermediates.

A) A Scheme showing the protocol has been exploited for neural differentiation of iPSC toward cortical neurons. B) Phase-contrast images of different stages of neural differentiation: early neural rosette (left), late neural rosettes (middle) following manual picking and plating of the early rosettes, and established cultures of neural progenitors (right). C) Immunostaining of early neural rosettes, late neural rosettes and established neural progenitors showing the expression of neural stem cell markers Nestin (red), Pax6 and Sox2 (green). Nuclei were stained with DAPI (blue). D) Western blot showing complete silencing of Oct4 protein, including no expression of transgene-encoded Oct4, in 2 independent lines of iPSC-derived neural progenitors. Undifferentiated iPSC expressed Oct4 and were used as controls (right lane). Vinculin was used as loading control. E) Phase contrast images of the cortical neurons post 70 days (left), neurons are infected by lentivirus expressing GFP under the promotor of

Synapsin. GFP positive neuron revealing the presence of synapsin expression in the mature neurons (middle), patient derived terminal neurons are positive for MAP2AB and βIII-tubulin (right).

4.4. Establishment of a large cohort of transgene-free induced pluripotent stem cell lines from WBS and 7dupASD patients

Although my results showed that virus transgene excision is not required to ensure robust differentiation of iPSC toward neural lineages, residual presence of the virus transgenes could still generate unexpected long-term instability and tumorigenicity due to permanent genetic integration(Pasi, Dereli-Oz et al. 2011), To tackle this problems, I successfully established the most innovative reprogramming technology that is based on the daily transfection of synthetic mRNAs encoding the five pluripotency factors OCT4 (also known as POU5F1), SOX2, KLF4, LIN28 and c-MYC (Warren, Manos et al. 2010) (Fig. 3.3A). Besides a higher efficiency in terms of number of reprogrammed colonies and kinetics, this integration-free approach avoids the residual permanence of reprogramming transgenes and its detrimental impact in terms of inter-clones heterogeneity, variability in differentiation proficiency, insertional mutagenesis and reprogramming factors-induced DNA damage. Epithelial mesenchymal transition during reprogramming was tracked by nuclear GFP mRNA expression included in the transfection cocktail (Fig. 3.3B). Two weeks post transfection, three iPSC colonies per line (26 lines in total) were selected and picked on the basis of their typical ES-like morphology and positive staining for Tra-1-60 live antibody (Fig. 3.3C).

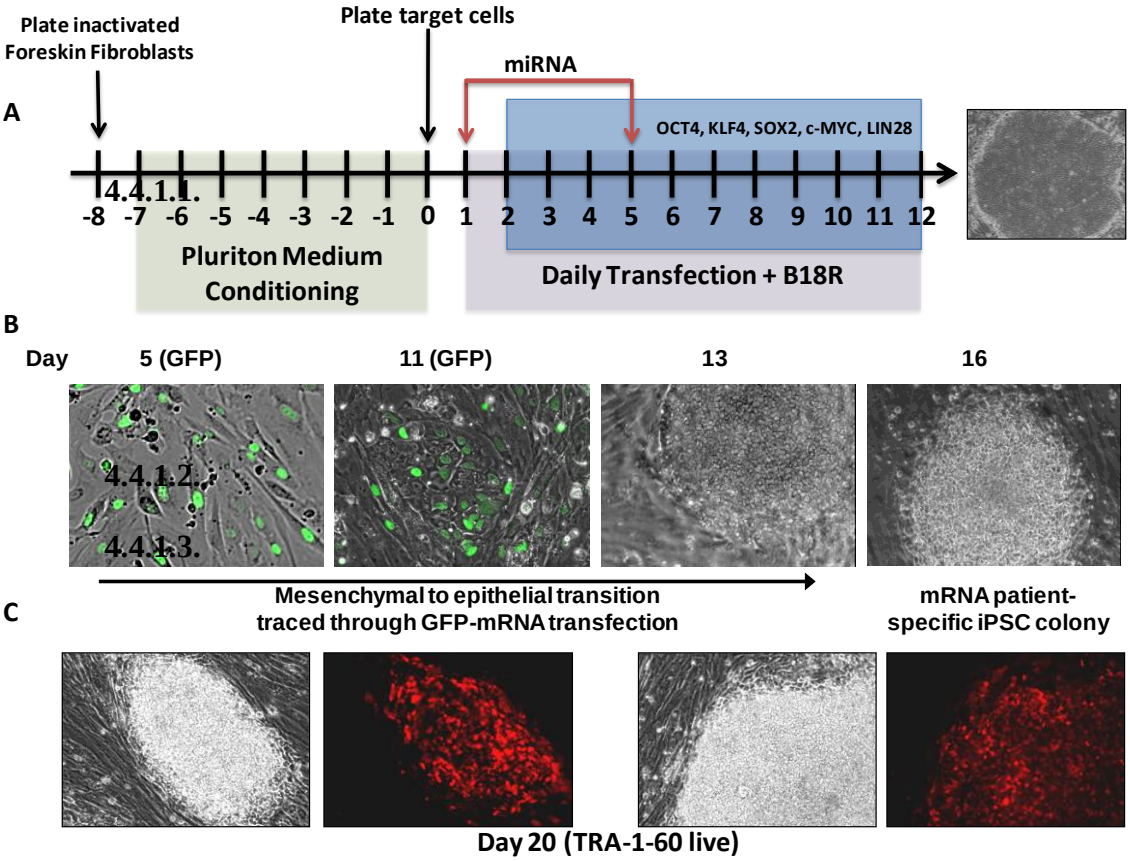


Fig 3.3 Establishment of a large cohort of transgene-free induced pluripotent stem cell lines from WBS and 7dupASD patients.

A) A scheme showing the various steps during mRNA based reprogramming. B) GFP tracking of mesenchymal-to-epithelial transition and C) Tra-1-60 live staining used to identify fully reprogrammed cells prior to picking.

I then characterized 3 independent iPSC lines from each patient or unaffected relative, along with 2 independent iPSC lines from the unrelated control individual, and one additional iPSC line previously reprogrammed by a conditional lentiviral vector following Cre-mediated excision of the single copy integrant amounting to a total cohort of 27 independent iPSC lines (Fig. 3.4 A).

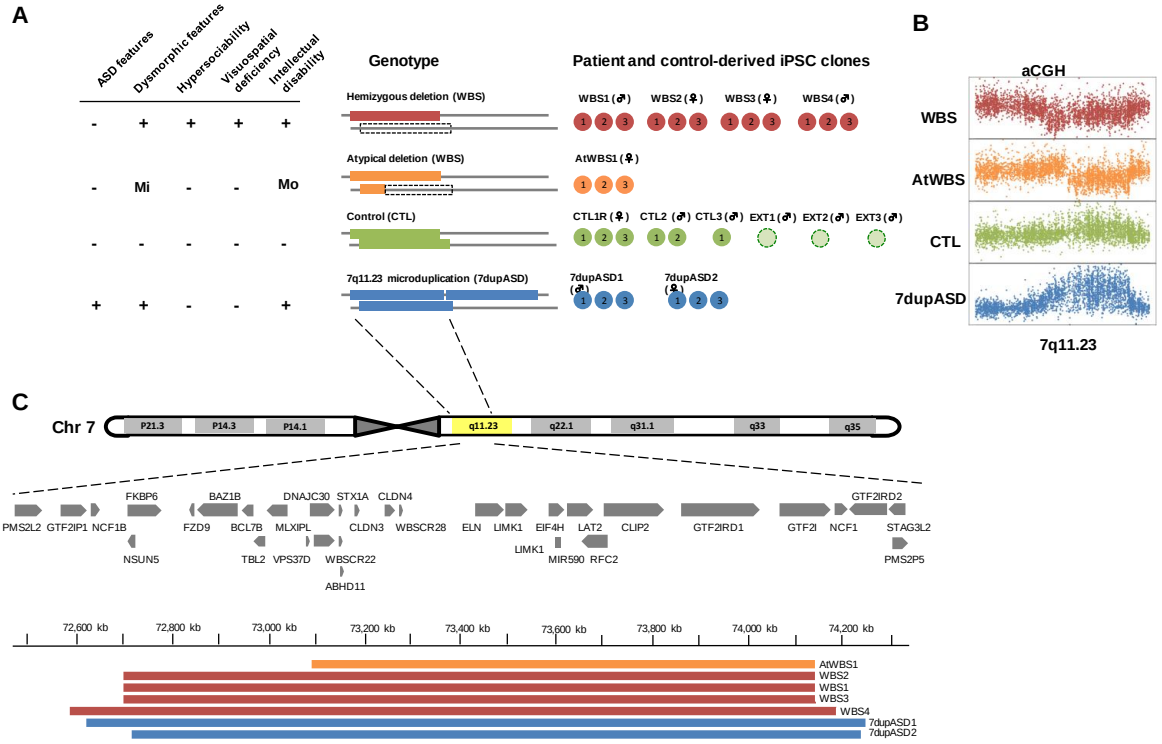


Fig.3.4 Scheme showing the summary of clinical information overlaid onto the configuration of 7q11.23 rearrangements in individuals recruited for this study.

A) Schematic representation of the WBS genetic interval (upper panel) and of the cohort of recruited patients including the number of independent iPSC clones derived per patients and a diagram showing the repertoire of clinical symptoms and cognitive behavioural traits (lower panels). Each genetic condition and the type of genetic rearrangement are represented with specific colors: typical WBS deletion (red), atypical deletion (orange) and 7q11.23 microduplications (blue). iPSC lines derived from healthy individual are also shown (green), as well as external controls added for differential expression analysis. B) Copy number of Array-CGH probes in the 7q11.23 region for a patient sample representative of each genetic condition. One representative control line is included. C) Schematic representation of the WBS genetic interval and boundaries of the CNVs detected by aCGH.

iPSC lines exhibited typical morphology, expressed the full range of pluripotency markers including ALP, OCT4 (POU5F1), NANOG, Tra-1-60, SOX2 and SSEA-4, (Fig 3.5 A-B) and contributed to the three embryonic germ layers upon teratoma assay in vivo (Fig. 3.5 A).

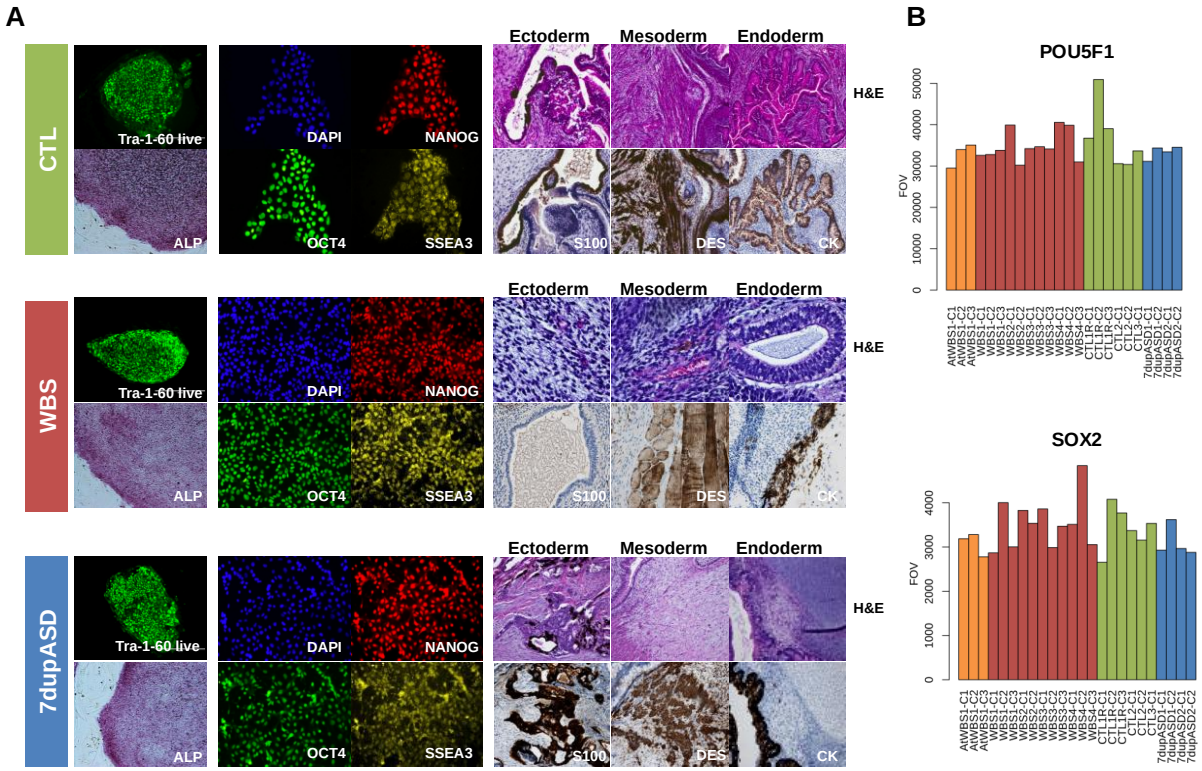


Fig 3.5. Characterization of mRNA based reprogrammed lines in this study

A) Expression of ALP by immunohistochemistry. Immunostaining for pluripotency markers (left). Nuclei were stained with DAPI (blue). Staining of three representative iPSC-derived teratomas showing expression of markers specific for the three germ layers (right). B) Nanostring mesurement of POU5F1 (also known as OCT4) and SOX2 expression in the panel derived iPSC lines. Level of expression is indicated as normalized fields of view (FOV).

4.5. Genomic instability of mRNA based reprogrammed iPSC

During the recent years, there have been enormous efforts in understanding of the mechanisms underlying the genomic integrity in ESC and iPSC as it is vital for the future use of such pluripotent cells in disease modeling, regenerative medicine and drug screening fields.

For examples, alterations at the genomic level not only can lead to failure of transplanted cell function, but more importantly could potentially lead to tumorigenesis. Moreover, such changes at DNA lever may influence the cell's response to drugs, thus compromising the fidelity of drug screens and counter screens(Grskovic, Javaherian et al. 2011).

In order to assess the genomic integrity of mRNA based reprogrammed iPSC lines, I isolated genomic DNA from at least 2 iPSC clones per individual along with their parental fibroblast and then we subjected them to high density CytoScan Arrays. As shown in Table 2, very few copy number variants were identified in iPSC lines, most of which already pre-existed in parental fibroblasts, consistent with recent evidence of pronounced CNV mosaicism in human skin from which ‘de novo’ iPSC-specific CNV emerge as a result of clonal expansion rather than genomic damage.

Backgrou	Cell Type	Sample	Number of CNVs		Potentially functional CNVs, excluding DGV and WBS region		Affected regions (IPSC-specific)
			All	IPSC-specific	All	IPSC-specific	
7dupASD1	Fibroblast		3		0		
7dupASD1	miPSC	C1	5	2	2	1	20q11.21 (gain)
7dupASD1	miPSC	C2	4	1	0	0	0
7dupASD1	miPSC	C3	5	2	2	1	6q16.3 (loss)
7dupASD2	Fibroblast		5		0		
7dupASD2	miPSC	C1	6	1	0	0	0
7dupASD2	miPSC	C2	20	15	14	14	14 ** *array fails quality controls*
7dupASD2	miPSC	C3	5	0	0	0	0
CTL1R	Fibroblast		3		0		
CTL1R	miPSC	C1	3	0	0	0	0
CTL1R	miPSC	C2	7	4	1	1	6p21.32 (gain)
CTL1R	miPSC	C3	9	6	5	5	1q43 (gain); 5q32 (gain); 5q12.3 (gain); 6q21 (gain); q23.3 (gain)
CTL2	miPSC	C1	7	NA	7	NA	(fibroblast not available)
CTL2	miPSC	C2	4	NA	4	NA	(fibroblast not available)
CTL3	viPSC	C1	4	NA	4	NA	(fibroblast not available)
WBS1	Fibroblast		4		0		
WBS1	miPSC	C2	6	2	0	0	0
WBS1	miPSC	C3	7	3	2	2	3p12.3 (gain); 7q31.32 (gain)
WBS2	Fibroblast		6		0		
WBS2	miPSC	C1	8	2	1	1	7q31.32 (gain)
WBS2	miPSC	C2	9	4	3	3	6q16.1 (gain); 7q21.11 (gain); 7q31.32 (gain)
WBS2	miPSC	C3	5	2	2	2	4q22.1 (gain); 7q31.32 (gain)
WBS3	Fibroblast		4		0		
WBS3	miPSC	C1	4	0	0	0	0
WBS3	miPSC	C2	9	5	3	1	7q31.32 (gain)
WBS3	miPSC	C3	7	3	2	2	6q16.1 (gain); 7q31.32 (gain)
WBS4	Fibroblast		4		0		
WBS4	miPSC	C1	7	3	0	0	0
WBS4	miPSC	C2	10	2	0	0	0
WBS4	miPSC	C3	8	4	3	3	6q16.1 (gain); 7q31.32 (gain); 6q21.1 (loss)
AtWBS1	Fibroblast		4		0		
AtWBS1	miPSC	C1	11	7	5	5	6q16.1 (gain); 7q21.11 (gain); 7q31.32 (gain); 20q13.33 (gain); 19q13.42 (gain)
AtWBS1	miPSC	C2	5	1	0	0	0
AtWBS1	miPSC	C3	11	7	5	5	1p31.1 (gain); 6q16.1 (gain); 7q21.11 (gain); 7q21.11 (gain); 19q13.42 (gain)

Tab. 3.2 Summary of the CNVs identified through aCGH.

4.6. Expression of 7q11.23 genes follows gene dosage in the pluripotent state

In order to ascertain whether the pluripotent state represented a meaningful stage at which to probe the effect of 7q11.23 dosage, we first asked whether the mRNA expression of the 7q11.23 genes follows gene dosage. For this we resorted to the high accuracy of Nanostring-based quantification as well as to RNAseq and found that the expression of all genes of the interval (including those expressed at very low levels) mirrors gene dosage (Fig. 3.6A), thus excluding compensatory effects from the wild type allele. We then confirmed that also at the protein level the expression of both GTF2I and BAZ1B, the genes associated to key traits of WBS and 7DupASD(Hirota, Matsuoka et al. 2003, Edelmann, Prosnitz et al. 2007, Lazebnik, Tussie-Luna et al. 2009, Antonell, Del Campo et al. 2010, Sakurai, Dorr et al. 2011, Malenfant, Liu et al. 2012), reflected the symmetrical dosage of the two conditions (Fig. 3.6 B-D).

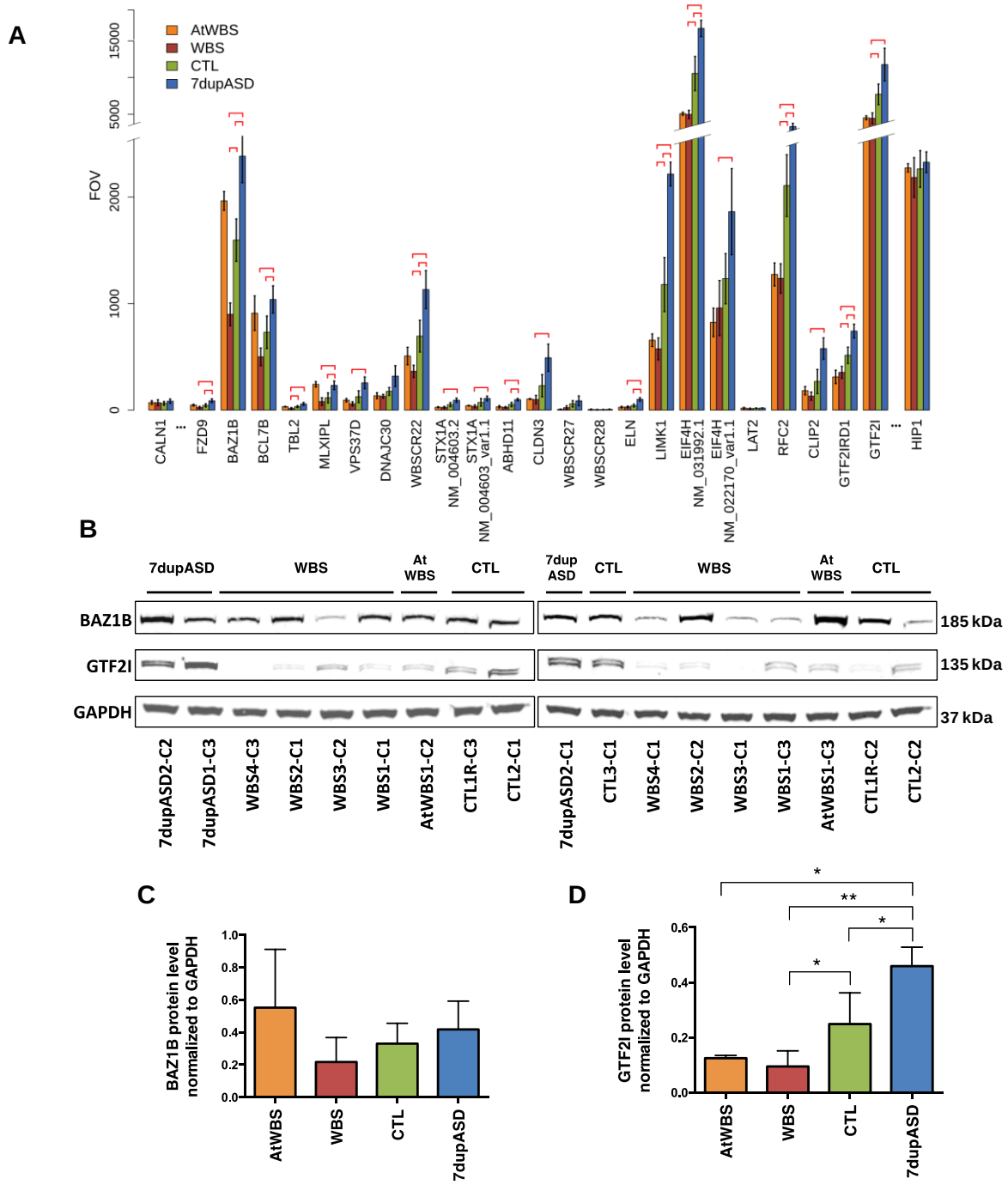


Fig. 3.6 Expression of 7q11.23 genes follows gene dosage in the pluripotent state

A) Nanostring quantification of the expression of genes included in the WBS genomic interval at the iPSC stage. Error bars represent the standard deviation in each genetic condition, while the horizontal bars above the respective comparisons indicate statistical significance. B) Western blot and (C-D) densitometry analysis of GTF2I and BAZ1B levels in a large portion

of WBS, 7dup-ASD and At-WBS samples compared to control (CTL). *: $p < 0.05$, **: $p < 0.01$.

4.7. 7q11.23 dosage imbalance causes transcriptional dysregulation in disease-relevant pathways already at the pluripotent state

To assess differential expression between genotypes, we profiled by RNAseq the panel of patient- and control- derived iPSC lines, and complemented this dataset also with additional control lines from the literature (hereafter referred to as external controls, see methods), excluding from further analysis the genes that were differentially-expressed between controls from our cohort and external controls. A pair-wise comparison of the three genotypes identified 757 differentially-expressed genes (DEGs) (Fig. 3.7A). Strikingly, Gene Ontology (GO) analysis of the union of DEGs revealed significant enrichments for biological processes of obvious relevance for the hallmark phenotypes and target organ systems of the two conditions. Fig. 3.7 B shows a treemap representation of the most specific enriched biological processes (the full list of GO enrichments for each comparison are in shown is Fig. 3.7 C-F), in which square sizes are proportional to the significance of the enrichment The top-ranking categories are related on the one hand to cell adhesion, migration and motility, which appear especially relevant in light of the wide range of connective tissue alterations that characterize WBS, and to the nervous system, providing a molecular context for the defining neurodevelopmental features of the two conditions. In addition, further enrichments relate to remarkably specific features of the two diseases which include: i) cellular calcium ion homeostasis, a category of potential relevance across disease areas but that acquires particular salience in light of the high prevalence of hypercalcemia in WBS(Kruse, Pankau et al. 1992); ii) inner ear morphogenesis, consistent with the combination of hyperacusis and sensorineural

hearing loss that is virtually always present in WBS(Gothelf, Farber et al. 2006), as well as with the balance and sensory processing disorders found in ASD(Kern, Trivedi et al. 2006); iii) a number of categories relevant for the craniofacial phenotype, such as skeletal muscle organ development, migration and neural crest cell differentiation; and iv) categories such as blood vessel development and cardiovascular system development, that reflect the wide range of cardiovascular problems in WBS; v) kidney epithelium development, in line with the highly prevalent kidney abnormalities of WBS(Pankau, Partsch et al. 1996). Importantly, removal of the external controls did not lead to significant changes in the enrichments we obtained (Fig. 3.7F), indicating that the cohort of our in-house reprogrammed lines already sufficed to capture the key features of 7q11.23-dependent transcriptional dysregulation. Furthermore, in order to exclude the possibility that such enrichments could arise by chance, we performed a series of shuffling tests entailing comparisons between randomly assigned groups of samples (see methods). In the rare cases in which these tests yielded any differentially expressed genes, the only significant GO enrichments arising within their union were related to extracellular matrix organization and response to mechanical stimulus. This points to these genes as particularly variable across lines and/or individuals, suggesting that enrichments related to these particular categories should be generally interpreted with care and underscoring the importance of subjecting GO enrichments to this rigorous scrutiny. Our analysis thus confirms that with the exception of these categories, the enrichments we found are specifically caused by 7q11.23 dosage imbalances.

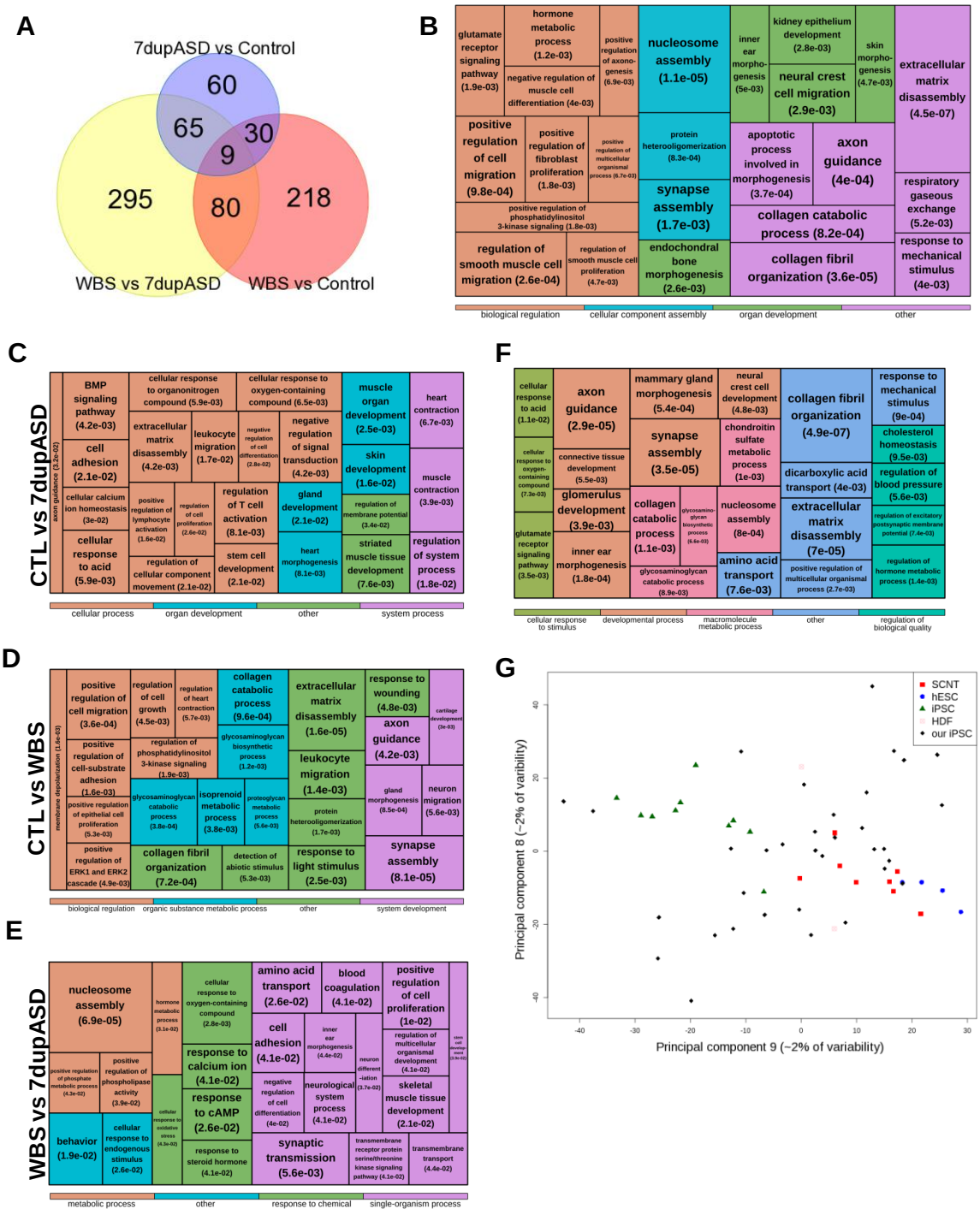


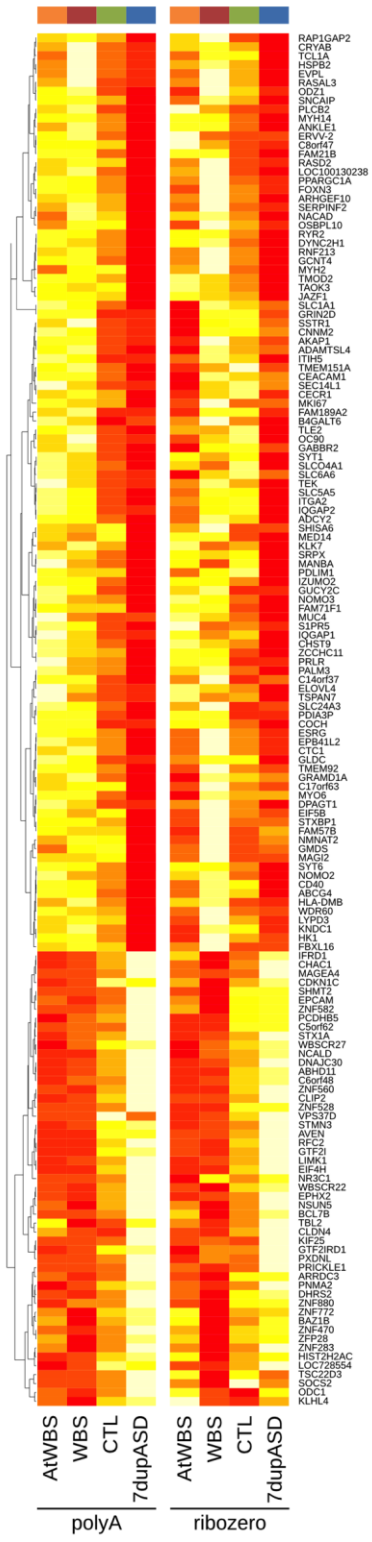
Fig. 3.7 Analysis of the transcriptomic changes caused by 7q11.13 CNVs.

A) Number and distribution of differentially-expressed genes (DEGs) among the three comparisons. B) Top most-specific enrichments for GO biological processes among DEGs. Parent categories with enriched children categories were filtered out; the color code indicates parent categories that have been selected approximating the best non-overlapping combination

of parents. DEGs show enrichment for categories recapitulating all aspects of the diseases. C-E) GO biological processes enriched among DEGs between Control *vs* 7dupASD**,** WBS *vs* Control and WBS *vs* 7dupASD. F) Enrichment for GO biological processes among the union of DEGs when excluding the external control lines from the analysis. G) Principal component analysis of the published data comparing iPSC, IVF-derived hESC and SCNT-ESC[16]. Plotted are the first components able to distinguish, in the published data (after trimmed mean of M-values normalization with ours), between iPSC and SCNT. Although our lines span the spectrum of variation on these components, most of them side with SCNT-ESc and IVF-ESc.

We found that the majority of DEGs either show a symmetrically opposite pattern in the two conditions or have a fold-change in the same direction over controls, indicating that the symmetrical dosage imbalance affect mostly the same transcriptional programs, either in the same or in symmetrically opposite ways. We thus proceeded to uncover what were the most highly symmetrical genes, taking the DEGs for which the mean expression in control samples was within a 20-80% range between the means of the WBS and 7DupASD (~39% of the DEGs) and that had an absolute Pearson correlation of at least 0.5 with WBS gene dosage.

This high-confidence set included 166 symmetrical DEGs (Fig. 3.8) (enriched for the single GO category of synaptic transmission), establishing that, in the pluripotent state, the symmetry in dosage is reflected into at least 22% of transcriptional dysregulation. Notably, this set includes genes associated to characterizing phenotypes of the 2 conditions, as in the case of PDLIM1 and MYH14, the former associated to attention-deficit disorder(Wang, Liu et al. 2012), neurite outgrowth(Ohno, Kato et al. 2009), cardiovascular defects, and hyperacusis, and the latter involved in hearing impairment and lip development(Martinelli, Arlotti et al. 2008).



4.8. Transcriptional dysregulation in disease relevant cell types

That 7q11.23 CNVs trigger disease-relevant transcriptional dysregulation already at the pluripotent state suggests that these initial conditions may prime the accumulation of even greater trascriptional alterations during development. Further, it predicts that the aggregate dysregulation in iPSC across categories spanning several developmental pathways and organ systems will be channeled upon differentiation, resulting in the selective amplification of specific domains of iPSC-specific dysregulation in a lineage-dependent fashion.

In order to test this hypothesis, we differentiated our iPSC lines into three lineages of cardinal relevance for the two conditions (Fig. 3.9): i) Pax6+ telencephalic neural stem cells and progenitors (NPC)(Shi, Kirwan et al. 2012, Shi, Kirwan et al. 2012); ii) neural crest stem cells (NCSC)(Menendez, Kulik et al. 2013), which originate the craniofacial structures along with several other disease-relevant lineages; iii) mesenchymal stem cells (MSC)(Menendez, Kulik et al. 2013), hierarchically upstream of osteocytes, chondrocytes, smooth muscle cells, and other physiopathologically meaningful cell types.

Fig. 3.8 Genes differentially expressed in a symmetrical manner in

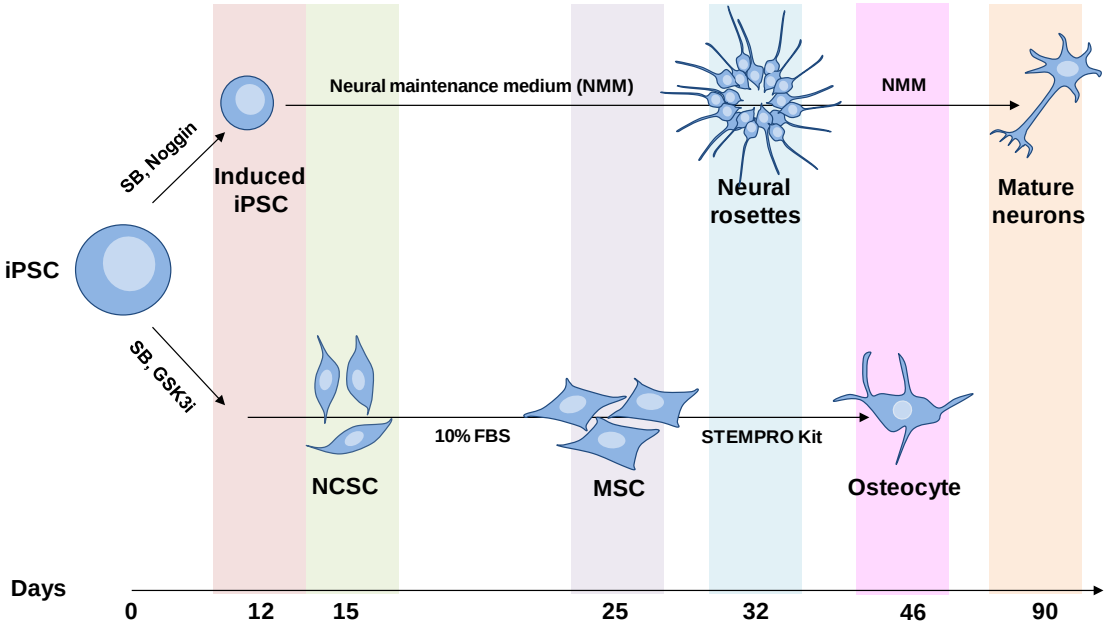
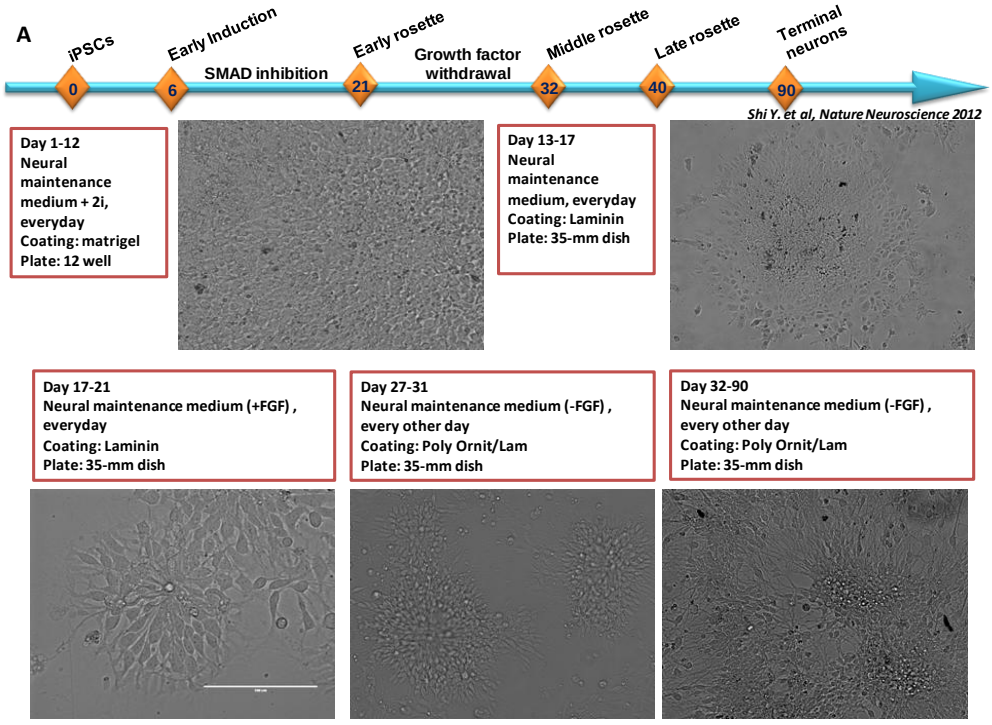


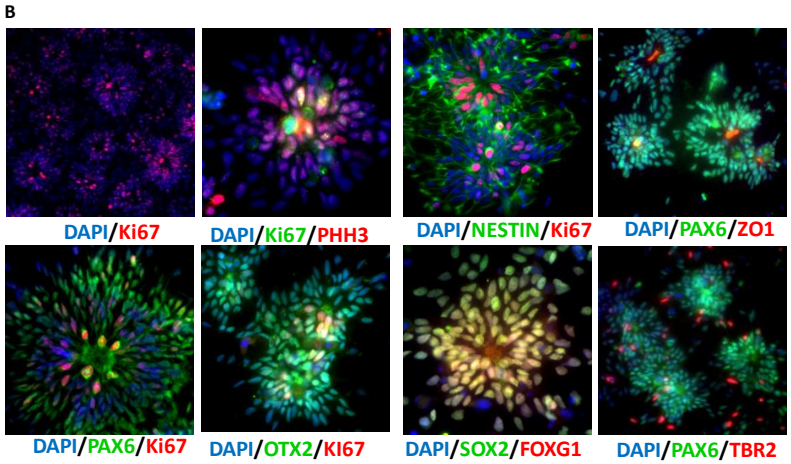
Fig. 3.9 Scheme of iPSCs differentiation protocols toward cortical neurons (above) or osteocytes (below).

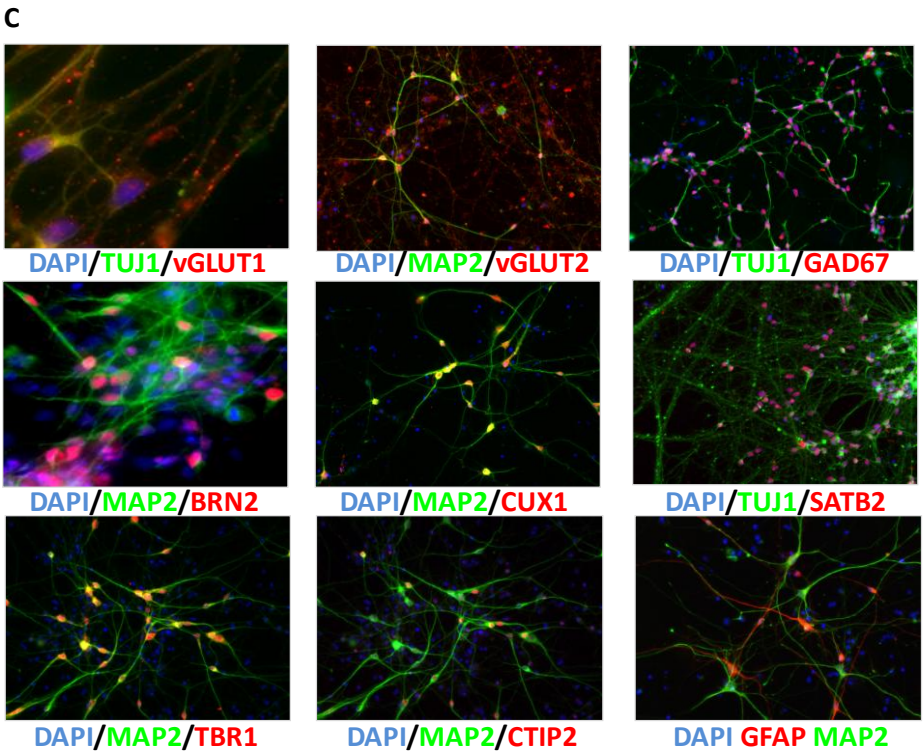
4.9. Differentiation of iPSC toward cortical neuron using small molecules

I successfully implemented the differentiation of iPSC into the dorsal telencephalic lineage by dual Smad inhibition in the presence of SB431542 and Noggin (Chambers, Fasano et al. 2009, Shi, Kirwan et al. 2012) (Fig. 3.10A). This protocol not only allows differentiation iPSC toward primary neuroepithelial cells but also captures in-vitro the three main progenitor populations of the human developing cortex (ventricular radial glia, intermediate progenitors and basal radial glia) which in turn generate the full range of early-born deep-layer and late-born upper-layer cortical neurons, thereby allowing the in vitro recapitulation of human corticogenesis (Shi, Kirwan et al. 2012, Shi, Kirwan et al. 2012) . I confirmed that patient-derived NPCs expressed key forebrain markers such as FOXG1 (telencephalic marker), OTX2 (forebrain marker) and ZO1 (a marker for tight junctions, expressed in the lumen of neural

rosettes), and were typically arranged in neural rosettes with TBR2+ intermediate progenitors surrounding apical Pax6+ progenitors (Fig.3.10B). Next, I directed the differentiation of neural progenitor cells toward terminally differentiated neurons by withdrawal of FGF. Cortical neurons then stained for the panel of lower and upper layers markers (Fig.3.10C) showing that this protocol is able to recapitulate key event of corticogenesis in-vitro. I profiled 15 patient and control-derived NPC lines and found that most of the differentially expressed genes were enriched in GO categories related to neuronal function such as axon guidance, regulation of transmitter secretion and negative regulation of axonogenesis (Fig. 3.10D).







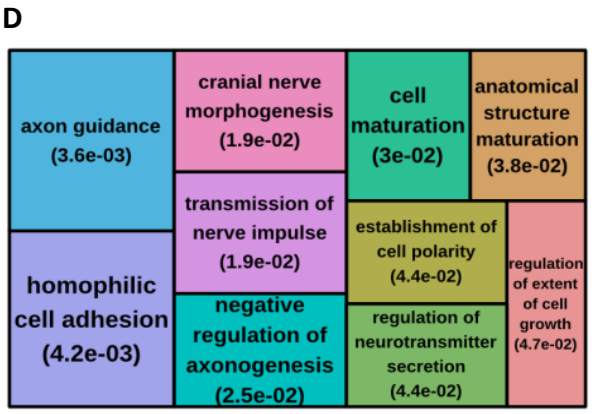


Fig. 3.10 Differentiation of iPSC toward cortical neuron using small molecules.

A) Scheme of Differentiation of iPSCs to cortical neurons through stem/progenitor cells. Generation of cortical glutamatergic neurons from human iPSCs *in vitro* recapitulates *in vivo* corticogenesis. B) iPSC-derived cortical stem/progenitor cells recapitulate the emergence of stem cell populations in human corticogenesis. Rosettes were stained for proliferating (Ki67), mitotic (phospho-histone H3) and neural stem cell (NESTIN, ZO1, and PAX6) markers (above). Default forebrain specification is evidenced by the expression of OTX2, FOXG1 and SOX2 markers (below). C) Images of human iPSC-derived cortical neurons expressing neuronal marker Tuj1. Deep and upper layer neurons were generated in the expected temporal

order, with layer 5 and 6 neurons (TBR1+, CTIP2+ and FOXP1+) emerging before layer 2/3 neurons (BRN2+). Upper layer, later born cortical neurons, defined by the expression of CUX1, SATB2 and BRN2 transcription factors, emerge in this system several weeks after the early born, deep layer neurons. The overwhelming majority of neurons eventually acquire glutamatergic fate, as shown by the widespread appearance of vGlut1 and vGlut2 punctae on neurites. D) Top enrichments for GO biological processes among NPC DEGs.

4.10. Differentiation of iPSC to neural crest stem cells

Craniofacial dysmorphic features are considered as one of the key hallmarks of WBS while they have also reported in 7dup-ASD patients (Merla et al., 2010; Pober, 2010; Somerville et al., 2005). This could point to the fact that aberrant expression of some residing genes in WBSCR may affect the common pathways involved in craniofacial phenotype. To test this hypothesis, I differentiated the same cohort of iPSC lines into NCSC that displayed distinct morphology (Fig. 3.11A), and stained homogeneously positively for HNK1 and NGFR, a defining combination of NCSC markers, in immunofluorescence (Fig. 3.11B), by flow cytometry (Fig. 3.11C) and qPCR (Fig.3.11D). Transcriptional profiling revealed differential expression in 364 genes (GO enrichments shown in Fig. 3.11E), including key genes linked to craniofacial dysmorphisms (ATP2C1, HHAT, LMNB1, MAPK8, PTCH1 and SATB2)(Dobreva, Chahrour et al. 2006, Singh, Yin et al. 2007, Cobourne, Xavier et al. 2009, Bonilla-Claudio, Wang et al. 2012, Dennis, Kurosaka et al. 2012, Metzis, Courtney et al. 2013, Kurosaka, Iulianella et al. 2014, Zhao, Qu et al. 2014) and RhoA Signalling/Signalling by Rho Family GTPases (WASF1, GFAP, ACTR2, STMN1, MAPK8, ARHGEF11 and PLXNA1)(Minoux and Rijli 2010, Phillips, Papoutsi et al. 2012). Importantly, small GTPase RhoA signalling has been recently showed to rescue SM-actin filament bundle formation of smooth muscle cells in a model of WBS(Ge, Ren et al. 2012).

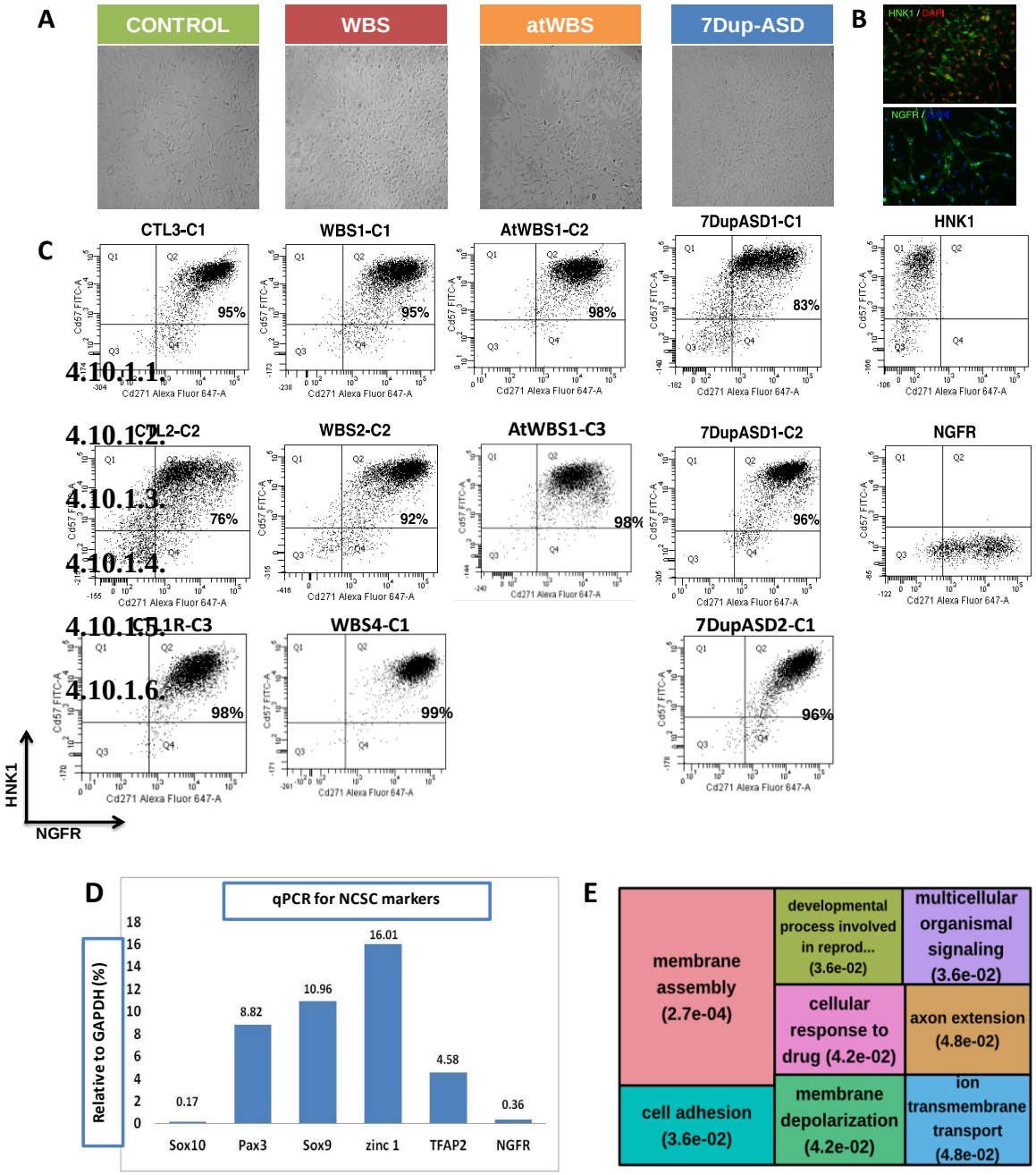


Fig. 3.11 Characterization of NCSC lines derived from WBS, atWBS, 7DupASD and control iPSC lines.

A) Phase contrast microscopy shows a similar morphology between the four genotypes. B) Immunofluorescence analysis indicates positivity for two NCSC markers (HNK1 and NGFR) in a representative iPSC-derived NCSC line. C) Flow cytometry analysis indicates a high

percentage of HNK1-NGFR double positive cells in NCSC. D) RT-qPCR showing the expression of key neural crest markers. E) Top enrichments for GO biological processes among NCSC DEGs.

4.11. Differentiation of iPSC toward mesenchymal stem cells

Finally, in order to define the impact of iPSC-primed transcriptional deregulation upon further differentiation, I induced patient-derived NCSC towards the MSC fate in the presence of fetal bovine serum (FBS). Already four days after induction, all cells gained a MSC-like morphology (Fig. 3.12A) and positively stained for key mesenchymal stem cell markers CD44 and CD73 (<95%) (Fig. 3.12B).

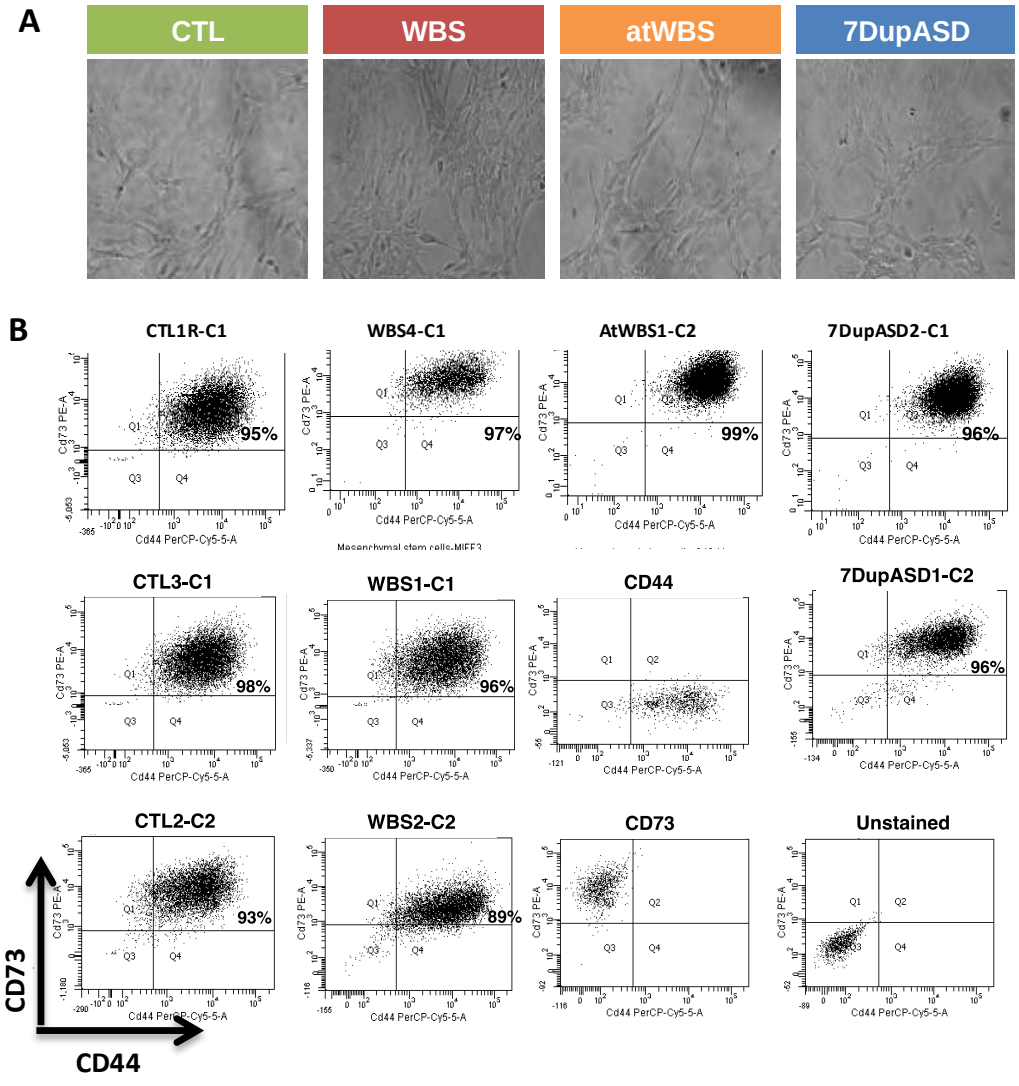


Fig. 3.12 Characterization of MSC lines derived from WBS, atWBS, 7DupASD and control iPSC lines.

A) Phase contrast microscopy shows a similar morphology between the four genotypes. B) Flow cytometry analysis of MSCs for CD73+ and CD44+ cells at day 10 of differentiation.

Transcriptomic profiling confirmed that, with the notable exception of ELN, also in MSC expression of 7q11.23 genes recapitulated dosage (Fig. 3.13 A-B), and yielded 422 DEGs showing enrichment for several GO categories related to tissue morphology (Fig. 3.13C). Interestingly, and in contrast to iPSC, NPC and NCSC, shuffling tests (see methods) yielded enrichments in many of these same categories, though most often with much lower enrichment and significance. This may be due to the discrete statistics used for the analysis of RNAseq (*vis* a *vis* microarray data), which might amplify the effect of variability across samples. Alternatively, it may suggest that the MSC state is less homogeneous than the pluripotent state, manifesting itself as a less narrow attractor that is hence more sensitive to inter-individual divergence in gene expression. Regardless, removal of these potentially spurious genes from the DEGs led to the identification of the same major categories enriched among DEGs, indicating that their significance does not hinge on these variable genes. In addition, some categories of great interest for the craniofacial phenotypes, such as apoptotic processes or skeletal muscle tissue development, were wholly undetected in the shuffling analysis. These findings confirm that, even against the backdrop of significant inter-individual variability in gene expression, 7q11.23 dosage significantly affects disease-relevant pathways. In fact, and even more remarkably, samples clustered by genotype at the whole transcriptome level (Fig. 3.13D), indicating that 7q11.23 dosage imbalances have an especially high penetrance at this developmental stage. Strikingly, the atypical sample (AtWBS1-C2)

clustered with the controls, suggesting that spared genes are particularly important in this lineage, in line with recent reports of the role of BAZ1B in neural crest migration(Barnett, Yazgan et al. 2012). Moreover, an Ingenuity Pathway Analysis (IPA) on DEGs between WBS and CTL revealed a molecular network (the highest ranking network) enriched for genes related to cardiovascular system development (Fig. 3.13E) and including key regulators of cardiovascular development (Fig. 3.13F).

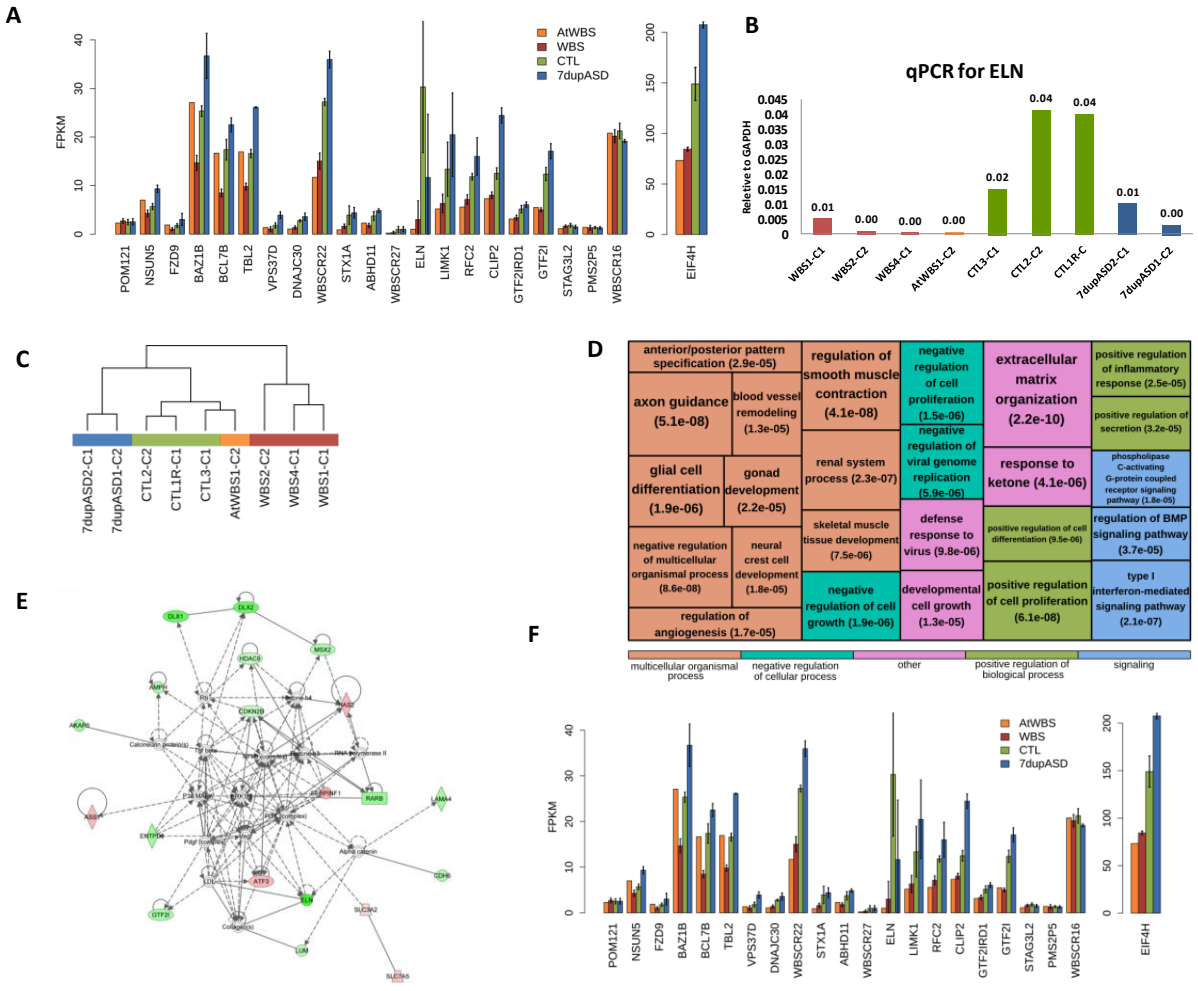
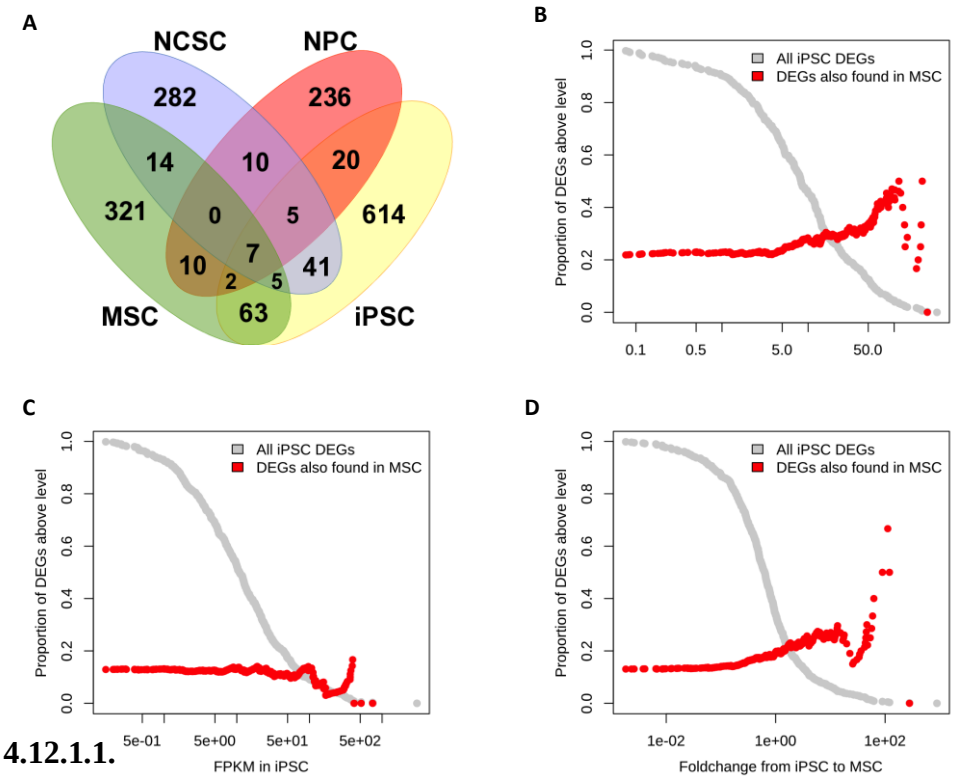


Fig. 3.13 Transcriptomic profile of MSC lines derived from WBS, atWBS, 7DupASD and control iPSC lines.

A) Plot of RNAseq expression levels of genes included in the WBS genomic interval at the MSCs stage. For better visualization, genes were separated into low/medium (right panel) and high expression (left panel). B) RT-qPCR showing the expression of key ELN among various genotypes. C) Unsupervised hierarchical clustering of correlations between MSCs whole transcriptomes, showing that samples cluster according to their genotype. D) Top most specific enrichments for biological processes among the MSC DEGs. E) Ingenuity Pathway Analysis on MSC DEGs reveals a molecular network enriched for cardiovascular system development. E) Expression of key members of the network in MSC. Error bars represent the standard deviation, while the horizontal bars represent statistical significance.

4.12. Statistically significant overlap of DEG between iPSC and disease relevant cell types

Given the physiopathological significance of the transcriptional dysregulation in MSC, we next asked what proportion of MSC-specific DEGs were also impacted at the iPSC state. As shown in Fig. 3.14A, 18% of MSC DEGs were already differentially-expressed in iPSC (25% when excluding external control iPSC), and the overlap steadily increases as we considered MSC DEGs with a higher expression in MSC, with 45% of the MSC DEGs expressed above 50 FPKM being found affected also in iPSC (Fig. 3.14B). Interestingly however, the proportion of overlapping DEGs did not correlate with expression at the iPSC stage (Fig. 3.14C), arguing against the hypothesis of greater accuracy at higher expression levels. We therefore hypothesized that genes that are dysregulated both in iPSC and MSC would be preferentially those that are specifically activated upon differentiation to MSC. Indeed, we found that of the iPSC DEGs that are down regulated upon differentiation to MSC (over 60%), very few remain differentially expressed also in MSC (Fig. 3.14D). In contrast, as we consider iPSC DEGs that increase expression upon differentiation to MSC, the proportion of DEGs maintained also in MSC rises to nearly 30%.



4.12.1.1.

Fig. 3.14 Comparison of DEGs among various lineages.

A) Overlap of DEGs identified in each lineage. Characterization of DEGs found in both iPSC and MSC. B) MSC DEGs that are also DEGs in iPSC have higher expression. C) The proportion of overlapping DEGs in MSCs does not correlate with expression levels in iPSC. D) The vast majority of DEGs at the iPSC stage is downregulated in differentiated MSCs and the overlap between iPSC and MSC DEGs increases with higher fold changes from iPSC to MSC.

4.13. IPSC-specific transcriptional dysregulation is amplified during development in a lineage-specific manner

On the basis of this analysis, we next hypothesized that the subset of iPSC DEGs that is conserved upon differentiation in each given lineage should be preferentially enriched in lineage-relevant categories. We thus evaluated, for each of the three differentiated lineages under study, the proportion of DEGs conserved for the GO categories that had been found enriched already in iPSC. As shown in Fig. 3.15 A-C and schematically represented in Fig.

4.1, upon differentiation iPSC DEGs are preferentially retained by category in a lineage-appropriate manner so that, for each target lineage, the proportion of conserved iPSC DEGs is much greater in categories relevant to that lineage (such as axonogenesis and axon guidance in the neural lineage, synapse-related categories in NCSC that will originate the peripheral nervous system and smooth muscle related categories in MSC).

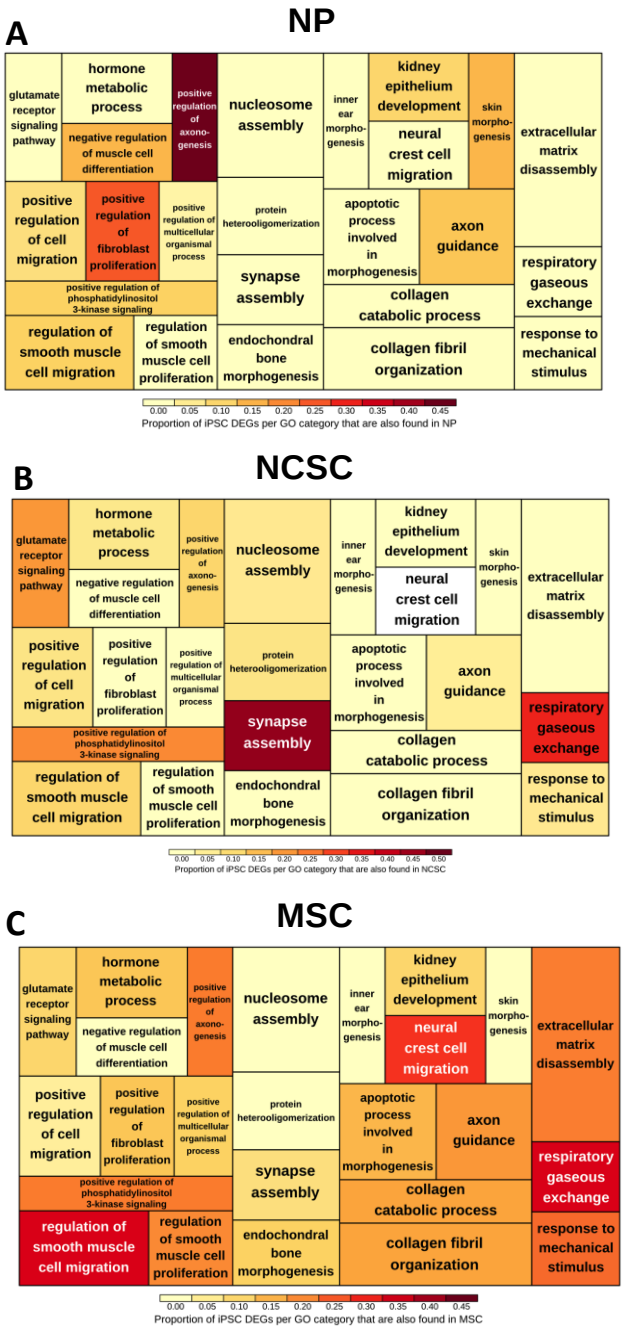


Fig. 3.15 Lineage-specific retention of iPSC DEGs.

(A-C) For each differentiated lineage, the treemap of enrichments that had been found among iPSC DEGs (Fig. 3.7B) is reproduced, plotting as a heatmap the proportion of iPSC DEGs in each category retained through differentiation.

Finally, we noted that the proportion of symmetrically dysregulated genes is significantly higher among the DEGs that are in common between iPSC and differentiated lineages (average odd ratio ~1.75, p~5e-03 evaluated from lineage-specific Chi-squared tests compounded through Fisher's method), supporting the notion that symmetrical patterns, likely under more direct control by 7q11.23 dosage, are particularly

relevant for the quota of disease-relevant transcriptional dysregulation that is seeded already in the pluripotent state.

4.14. A web platform for 7q11.23 CNV syndromes

Finally, the data I presented here were assembled by designing a new web platform called WikiWilliams-7q11GeneBase to make our data accessible to the community of scientists and practitioners working on these two diseases, through a user-friendly, gene-centered interface. Besides integrating, in a multi-layered manner, all our results with data from the literature, WikiWilliams is open to contributions by other groups through submission to the database's curators, with the aim of assembling in one site all molecular data on 7q11.23 syndromes (Fig. 3.16). The platform is openly accessible at http://bio.ieo.eu/wbs/.

Fig. 3.16 The WikiWilliams/7q11GB web platform.

A)Schematic representation of the data gathered in the open-access WikiWilliams/7qGBb web platform. B-D)Representative screenshot of the WikiWilliams/7q11GB database as it appears to users searching for a specific gene of interest. All transcriptomic and genomic data presented in this paper as well as previously published datasets can be easily interrogated in a multi-layered format integrated with several biological databases.

5. Discussion

In spite of increasing amount of information and huge investments by the pharmaceutical industry on genetic and age-associated disorders, the development of novel therapeutics in particular for cardiovascular abnormalities and neurological disorders has proven to be challenging and thus few new therapeutic compounds are presently entering the market. This is partially owing to experimental tools including in vitro cultures and animal models which can recapitulate only some of the specific traits of human disease and thus makes the modeling of human diseases in laboratory very difficult (Bellin, Marchetto et al. 2012).

The generation of induced pluripotent stem cells (iPSC) from terminally differentiated cells holds great promise for disease modeling, drug discovery and regenerative medicine (Ming et al., 2011). Ectopic expression of defined pluripotency factors generates patient-specific pluripotent cells that can be then differentiated into relevant cell types to gain mechanistic insights into disease physiopathology. This technology is enabling researchers to undertake studies for functional annotation of human genomes and modeling diseases 'in a dish', which was previously inconceivable, promising to align well defined genetic lesions to clinical data through molecular phenotypes in vitro (Bellin, Marchetto et al. 2012).

To this end, the critical challenge is twofold: i) define the extent to which early developmental lineages are informative about disease-relevant pathways affected by genetic mutations and, ii) assess the feasibility of reliably identifying those pathways beyond the sources of variability inherent to the iPSC-based approach (Cahan and Daley 2013).

Here I addressed these questions by focusing on a paradigmatic pair of genetic syndromes caused by symmetrical copy number variations (CNV) at 7q11.23: Williams-Beuren syndrome and Williams-Beuren region duplication syndrome that includes autistic spectrum disorder (7dupASD) (Sanders, Ercan-Sencicek et al. 2011). WBS and 7dupASD involve, respectively, the loss or gain of 26-28 genes and have a prevalence of between 1 in 7,500 and 1 in 10,000 (Somerville, Mervis et al. 2005, Pober 2010). WBS is characterized by cardiovascular symptoms and facial dysmorphism, along with the hallmark behavioral-cognitive profile that combines hypersociability with comparatively well-preserved language abilities, but severely compromised visuo-spatial processing, counting and planning (Merla, Brunetti-Pierri et al. 2010, Pober 2010). 7dupASD, in contrast, features varying degrees of ASD ranging from severe speech impairment to full blown autism, along with craniofacial dysmorphisms, among which some are similar and some symmetrically opposite to those of WBS patients (Somerville, Mervis et al. 2005, Van der Aa, Rooms et al. 2009). Finally, both syndromes are associated with anxiety and attention deficit hyperactivity disorder (ADHD). Thus, the two conditions are paradigmatic of a fundamental aspect of CNV-based disease pairs, namely the fact that symmetrically opposite CNV result in shared as well as symmetrical phenotypes. Yet, despite significant insight from mouse models (Osborne 2010, O'Leary and Osborne 2011, Campuzano, Segura-Puimedon et al. 2012, Mervis, Dida et al. 2012), the molecular pathways specifically affected by 7q11.23 CNV in the human lineages that are most relevant for disease phenotypes are yet to be uncovered.

Here we present the largest cohort of WBS and 7dupASD iPSC lines and differentiated lineages, in which we find that 7q11.23 dosage impacts disease-relevant transcriptional programs already in the pluripotent state. These alterations are partitioned into shared and

symmetrically opposite ones and are further exacerbated upon differentiation into disease-relevant lineages.

Somatic cell reprogramming is limited by: i) low efficiency of iPSC derivation; ii) the risk of insertional mutagenesis and iii) residual transgene expression from integrated vectors (Anokye-Danso et al., 2011; Pasi et al., 2011b; Takahashi et al., 2007). To tackle these issues, many non-integrating platforms have been developed which allows reprogramming of human somatic cells to iPSC using modified mRNA or specific clusters of miRNA (Warren, Manos et al. 2010, Anokye-Danso, Trivedi et al. 2012). Here for the first time, I unveil the combinatorial potential of the reprogramming approach through the generation and full characterization of iPSC from a large cohort of patients affected by WBS and 7Dup-ASD based on miRNA enhanced-mRNA method in feeder-free conditions. I validated the pluripotency of these cells on the basis of expression of well-recognized pluripotency markers, maintenance of undifferentiated morphology upon several passages in feeder-free conditions and tri-lineage and lineage-directed differentiation in-vitro and in-vivo. Moreover, genomic integrity of iPSC lines were assessed at high resolution, finding only few CNV, the overwhelming majority of which pre-existed in parental cells, consistent with recent studies showing CNV mosaicism within parental cells prior to reprogramming (Abyzov et al., 2012). These results showed that the combination of miRNA and mRNA based approaches through a novel RNA delivery technology, besides preventing insertional mutagenesis and the persistent expression of reprogramming factors inherent to viral-based technologies, offers unprecedented efficiency. Furthermore, the addition of miRNAs to the previously described mRNA cocktail (OCT4, SOX2, KLF4, c-MYC, LIN28) further shortens the process to less than two weeks, reducing the time needed to obtain clinically relevant integration-free human iPSC and without detectable accumulation of reprogramming-induced mutations.

The unique size of this cohort, both in terms of number of individuals and of independent iPSC clones derived from each individual, allowed us to address two main questions I have mentioned above; first, to define the extent to which early developmental lineages are informative about disease-relevant pathways affected by 7q11.23 dosage imbalances, we looked at the expression of the WBS genes using high accuracy Nanostring-based quantification. The results showed that the expression of all WEB genes follows gene dosage. Next, the impact of such 7q11.23 dosage imbalances was probed by high throughput RNA sequencing. We found that 7q11.23 causes profound transcriptional dysregulation already in the pluripotent state, strikingly affecting major disease-relevant pathways. Although recently published gene expression profiling experiment in non-relevant cells types such as lymphoblasts have revealed abnormal regulation of gene pathways potentially related to relevant aspects of the WBS syndrome phenotype, (Antonell et al., 2010b), this is the first time that high-throughput transcriptional profiling at very early stage of development i.e., iPSC could detect advanced phenotype of disease such as cellular calcium ion homeostasis, inner ear morphogenesis, craniofacial phenotype, blood vessel development and cardiovascular system development and kidney epithelium development. This implies that physiological development is biased already from early embryogenesis as a result of 7q11.23 imbalances.

These findings also allowed us to answer the second main question in the field which is the feasibility of reliably identifying indentified pathways beyond the sources of variability inherent to the iPSC-based approach. As extensively discussed in the introduction , variability has been recently recognized as a key concern for iPSC-based disease modeling, inviting caution in the interpretation of results from few lines that do not adequately sample variability

either across individuals or across lines reprogrammed from the same individual (Cahan and Daley 2013).

In order to exclude the possibility that such enrichments in disease relevant pathway could arise by chance or as a result of variability in iPSC lines, we performed a series of experiment. First, we excluded the fact that such enrichment could be due to the fewer number of the control lines with respect to the disease derived iPSC line by including additional control from the literature. In order to minimize the potential bias produced by including external samples for only one of the conditions, we chose to add three samples (each coming from a different individual) -- that is, half the number of control lines we already had -- and excluded from further analysis any gene that was differentially-expressed between virus and mRNA-reprogrammed control lines. Although the vast majority of DEGs identified were already detected in the previous analysis, the introduction of new samples increased, as expected, variability among the controls and led to the loss of over a hundred DEGs (including some WBS genes whose differential expression we are very confident about based on independent Nanostring validation, thus highlighting inherent limitations in sensitivity when analyzing large cohorts of heterogeneous human samples by high throughput methods). Nevertheless, the main messages of the results remained entirely corroborated, and in fact some of the GO enrichments even increased in statistical significance. Importantly, the vast majority of differentially-expressed genes identified in this study were still found when removing the external controls, and most of the remaining DEGs were close to significance, arguing against the introduction of a major bias through the use of external controls.

In order to assess the effect, on the transcriptional analysis, of having fewer samples, we repeated the analysis of our polyA dataset (focusing on the comparison for which we had

the most samples, i.e. the global analysis of WBS *vs* CTL iPSC), using only subsets of the samples. Random removal of 1 clone per patient lead to a dramatic reduction in the number of DEGs (48 to 76% lost), and to the identification of DEGs that are falsified by the discarded data. The impact of removing all clones from one patient per condition (amounting to fewer samples than removing one clone per patient) was even greater. In contrast, depth of sequencing appeared to make little difference: reducing coverage by half led to the loss of 11% of DEGs and to very few false positives.

Next, to assess the possibility that the observed differential expression might arise due to random variations, we performed a series of differential expression analysis between randomly-selected samples, discarding comparisons in which the two groups were not balanced for sex and/or genotype. A minimum of 3 such combinations were tested per tissue, and the resulting genes were pooled for the purpose of enrichment analysis.

We first randomly assigned all patients to two groups, but we could obtain statistically significant genes in none of the combinations. We therefore gradually removed patients until significant genes were obtained, which did not happen until a comparison involving 6 vs 6 samples (in each group, 3 samples from 2 patients). In contrast, when clones were selected and assigned to groups in a way that maximized the number of patients represented in each group, we had to go down to 3 vs 3 (3 samples per group, coming from 3 different patients) to get statistically significant genes (18 DEGs, showing no significant GO enrichment). Similarly, random allocation of the control samples (including external controls, balanced across groups) yielded very few DEGs (maximum 17) and no significant GO enrichment.

These results suggest that the primary source of “spurious” differential expression is genetic variation between individuals, which only gets mitigated using lines derived from

several patients and thus confirms that with the exception of these categories, the enrichments we found are specifically caused by 7q11.23 dosage imbalances.

And finally, we confirmed that our iPSC lines share the pluripotent signature of published datasets and hence GO enrichment in disease relevant pathways could not be due to the spontaneous differentiation potential of our iPSC lines in culture. Indeed, with respect to the recent comparison of traditionally reprogrammed iPSC versus embryonic stem cell lines derived by nuclear transfer (NT-ES) or in vitro fertilization (IVF-ES)(Ma, Morey et al. 2014) the majority of our lines (over 2/3) sided with NT-ES and IVF-ES with respect to the transcriptional components distinguishing published iPSC from NT-ES/IVF-ES (Supplementary Fig. 3a).

Thus, by presenting the largest cohort of iPSC lines characterized so far for any single genetic condition, combined to the first large scale use of mRNA-based integration-free reprogramming, our study benchmarks the possibility of detecting robust dosage-dependent alterations in transcriptional programs, even when these are caused by subtle dosage imbalances.

We further confirmed this finding by differentiating our cohort of iPSC lines into three lineages of cardinal relevance for the two conditions: dorsal telencephalic progenitors and neural crest stem cells along with their further differentiated mesenchymal derivatives, respectively relevant for the cognitive/behavioral and craniofacial phenotypes of WBS and 7dupASD. We found not only that each differentiated cell type displayed specific transcriptional alterations that match the relevant disease domains, but also that these alterations were seeded to a significant extent already in iPSC and were in fact amplified, upon differentiation, in a lineage-specific manner.

The size and quality of our iPSC cohort not only permitted us to understand the impact of 7q11.23 CNV on transcriptional dysregulation in the pluripotent state; but more importantly allowed us to found out that this dysregulation was selectively amplified in a lineage-specific manner, with disease-relevant pathways preferentially and progressively more affected in differentiated lineages matching specific disease domains. The significance of this observation for the iPSC modeling field lies in the fact that the pluripotent state is by far the best characterized and most standardized one among the human developmental stages captured in vitro. Importantly, it is also the most amenable to high-throughput upscaling. Hence, the observation that the pluripotent state is not only a viable stage in which to measure disease-relevant transcriptional effects of genetic alterations, but that these effects are also predictive of further dysregulation in differentiated lineages, grounds the feasibility of middle-to-high-throughput iPSCs characterization in order to functionally annotate human genomes, prior to selecting lines and assays for more labor-intensive differentiation courses.

In terms of the molecular pathogenesis of WBS and 7dupASD, besides uncovering the impact of 7q11.23 dosage already in the pluripotent state, these results also provide a first entry point for the molecular dissection of the outstanding feature that characterizes these two conditions, namely the coexistence, in the face of symmetrically opposite CNV, of both shared and symmetrically opposite phenotypes. By analyzing many samples from both conditions, we were in fact able to define a subset of DEGs that follows a symmetrically opposite dosage-dependent trend. Importantly, we found that this quota is significantly retained upon differentiation, indicating that symmetrically opposite patterns of gene expression seeded already in the pluripotent state, likely under direct control of 7q11.23 dosage, become increasingly prominent in disease-relevant differentiated lineages, thus providing a strong rationale for studying these two diseases (and by implication other CNV-based symmetric

disease pairs) together. Importantly, our analysis of symmetrically dsyregulated targets also uncovered the following genes as prime candidates for mediating the molecular pathogenesis of defining aspects of the two conditions: i) PDLIM1, which has been associated to attention-deficit disorder, neurite outgrowth, cardiovascular defects, and hyperacusis; ii) MYH14, which was involved in hearing impairment and lip development; iii) BEND4, a TF harboring the BEN domain that distinguishes a recently characterized family of neural repressors.

Finally, the data I presented here were provided in a user-friendly, open source web platform in which we assembled the multi-layered datasets from this first cohort of WBS and 7dupASD samples, and which was designed to integrate ongoing contributions from the entire scientific community working on these two diseases, thus serving also as a first template for data sharing from iPSC-based functional genome annotation.

In conclusion, reprogramming the paradigmatic pair of genetic syndromes caused by symmetrical 7q11.23 CNV enabled us to define how gene dosage impacts disease-relevant pathways in early developmental lineages and disease relevent cell types. These findings have broad significance for the molecular pathogenesis of WBS and 7dupASD as well as for the reprogramming-based disease modeling field as a whole.

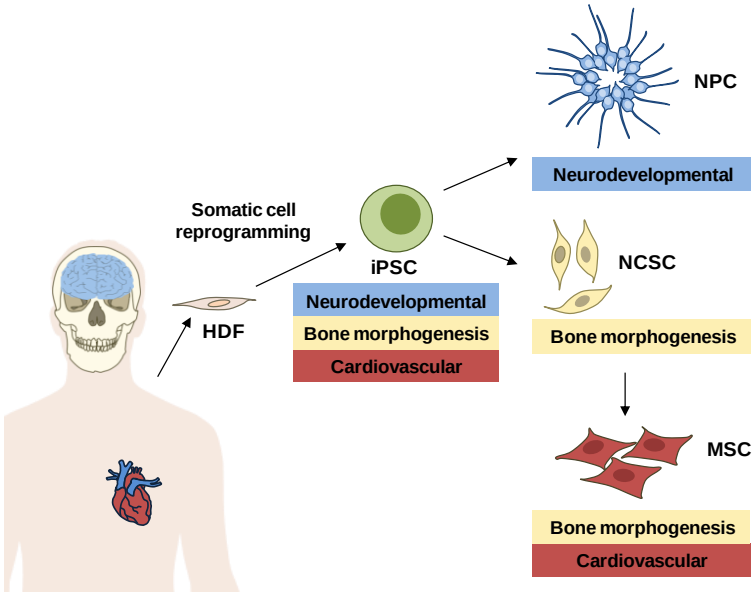


Fig.4.1 Graphical representation of the lineage specific retention of DEGs.

6. Future direction

6.1. A novel strategy for differentiation of iPSC toward homogenous population of cortical neural progenitors and neurons

Understanding the molecular mechanism of disease through somatic cell reprogramming riles to a large extent on the robust and reliable differentiation methods which can faithfully produce patient derived disease-affected cell types as pathophysiologically relevant in vitro models. To this aim, numerous protocols has been developed which enable differentiation of PSC toward neural and non-neural differentiation (Chambers, Fasano et al. 2009, Brennand, Simone et al. 2011, Shi, Kirwan et al. 2012, Menendez, Kulik et al. 2013) . Such a protocol should be able to produce step wise synchronized cells types in terms of differentiation which would allow us to understand the stages from which disease phenotype start to emerge and/or stages that mostly affected during differentiation process. Yet, development of a protocol which can synchronously produced NPC or cortical neuron is not fully addressed. To this goal and based on such considerations, I have designed and developed a protocol which can allow the selection of NPC cells which are positively expressing forebrain marker FOXG1. This protocol takes advantage of lentivrial vector in which puromycin resistance gene is expressed under the control of FOXG1 promoter and eGFP under control Of CMV promoter (Fig. 5.1A). Briefly, virus particles were produced using third generation packaging system in HEK 293 cells (Fig. 5.1B) and then iPSC cells were infected with the various concentrations of the virus and two day post infection FACS analysis performed to find the concentration of the virus which can maximize the infection efficiency

(Fig. 5.1 C-D). The infected iPSC which expressed GFP will be differentiated by the blockade of Smad signaling pathway (please see the materials and methods) and at the NPC stage, cells will be supplemented with puromycine to undergo selection. FOXG1 expressing cells will be assessed by immunostaing to ensure the identity of the selected NPC and prior to further differentiation. Finally NPC and differentiated cells will be compared with same cell line which has not undergone the selection steps and the results will be assessed by various technique including immunostaining and RNA-Seq.

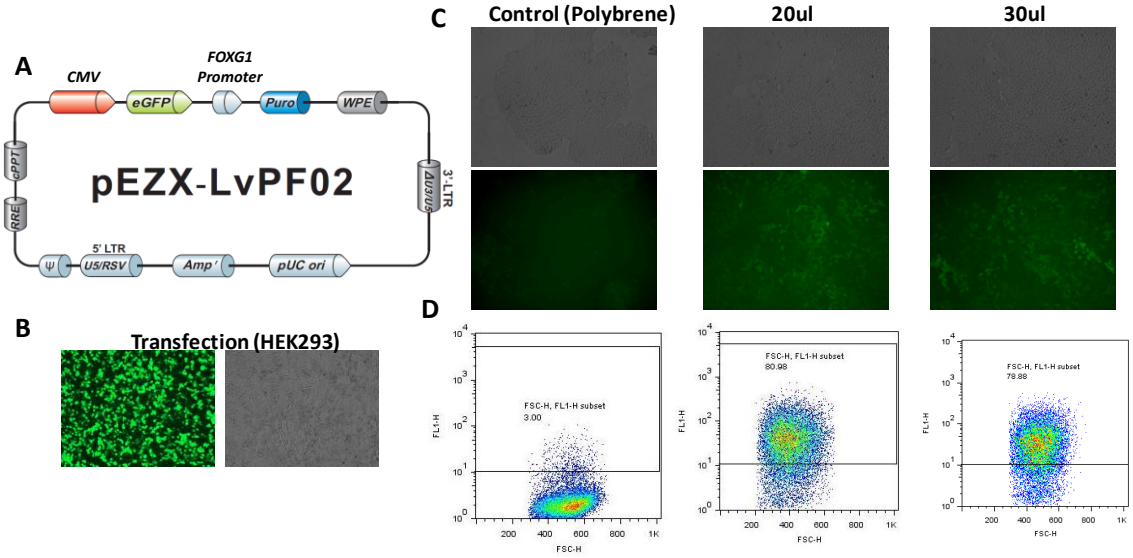


Fig. 5.1 A novel strategy for differentiation of iPSC toward homogenous population of cortical neural progenitors.

A) Scheme of the lentiviral vector with eGFP under control Of CMV promoter and puromycin under control of human FOXG1B promoter. B) HEK 293 cells expressed GFP one day post transfection. C) Optimizing viral infection to maximize the infection efficiency prior to differentiation. iPSC were infected with various concentration of the virus. D) FACS results showed the number of the infected cells in each condition.

6.2. Differentiation of homogenous population of MSC toward disease affected cell types

The second key question that I would like to address is to find a differentiation stage throughout MSC differentiation protocol in which various patients and healthy individual derived cell type are highly homogenous and thus they can be subjected for the further cellular or molecular characterization to probe the disease phenotype. My results showed that not all iPSC cells could give rise to homogenous population of the NCSC based on the expression of the two key NCSC markers, HNK1 and NGFR shown by FACS analysis (Fig.3.12). For this reason only, iPSC derived NSCS line which highly expressed these markers were selected for further differentiation toward MSC (Fig.3.11 and Fig.3.12). The unsupervised clustering analysis based on the global genes expression profiling showed that MSC but not NCSC were segregated based on their biological condition suggesting that MSC differentiation protocol is able to robustly differentiate NCSC toward highly homogeneous population of MSC. To further test this hypothesis, I perform another round of differentiation in which I subjected not fully differentiated NSCS lines along with not fully differentiated ones to MSC differentiation assay and asked if the differentiation media can favor the survival and maintenance of fully differentiated MSC. Strikingly, the FACS analysis showed that even from not HNK1 and NGFR double positive NCSC we can obtain a homogeneous population of the MSC which can express MSC key markers (Fig.5.2 A-B). In order to understand if the newly generated MSC lines will be clustered with their relevant genotype we will perform RNA-seq on these lines. Finally MSC will be further differentiated toward cell types which are affected in the patient such as smooth muscle cells and osteblasts in order to probe the disease phenotype in more advance stages of the disease.

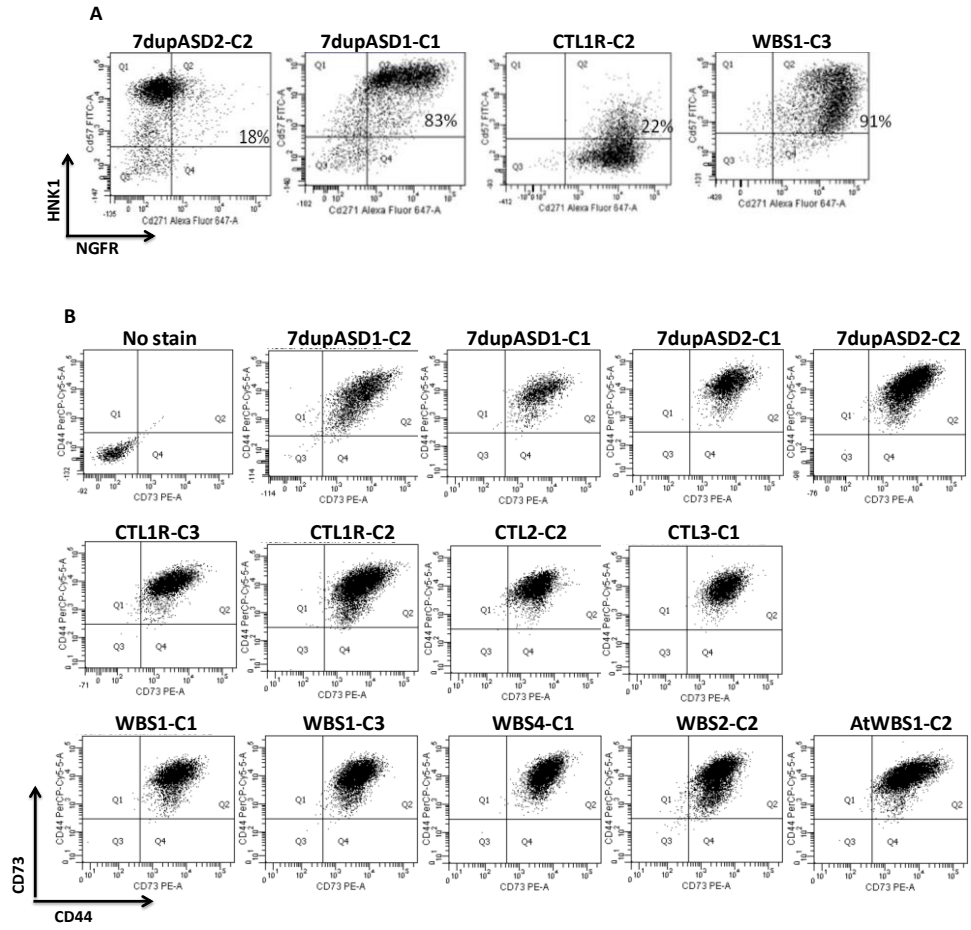


Fig. 5.2 Differentiation of NSCS toward MSC.

A) FACS analysis for iPCS derived NCSC lines showed that not all iPSC can efficeintly differentiate toward NCSC. B) FACS analysis for CD44 and CD73, two MSC key markers showed that even not fully differentiated NCSC can differentiate toward double positive MSC.

7. References

An, M. C., N. Zhang, G. Scott, D. Montoro, T. Wittkop, S. Mooney, S. Melov and L. M. Ellerby (2012). "Genetic correction of Huntington's disease phenotypes in induced pluripotent stem cells." Cell Stem Cell **11**(2): 253-263.

Anokye-Danso, F., C. M. Trivedi, D. Juhr, M. Gupta, Z. Cui, Y. Tian, Y. Z. Zhang, W. L. Yang, P. J. Gruber, J. A. Epstein and E. E. Morrisey (2012). "Highly Efficient miRNA-Mediated Reprogramming of Mouse and Human Somatic Cells to Pluripotency (vol 8, pg 376, 2011)." Cell Stem Cell **11**(6): 853-853.

Antonell, A., M. Del Campo, L. F. Magano, L. Kaufmann, J. M. de la Iglesia, F. Gallastegui, R. Flores, U. Schweigmann, C. Fauth, D. Kotzot and L. A. Perez-Jurado (2010). "Partial 7q11.23 deletions further implicate GTF2I and GTF2IRD1 as the main genes responsible for the Williams-Beuren syndrome neurocognitive profile." J Med Genet **47**(5): 312-320.

Antonell, A., M. Vilardell and L. A. P. Jurado (2010). "Transcriptome profile in Williams-Beuren syndrome lymphoblast cells reveals gene pathways implicated in glucose intolerance and visuospatial construction deficits." Human Genetics **128**(1): 27-37.

Ashe, A., D. K. Morgan, N. C. Whitelaw, T. J. Bruxner, N. K. Vickaryous, L. L. Cox, N. C. Butterfield, C. Wicking, M. E. Blewitt, S. J. Wilkins, G. J. Anderson, T. C. Cox and E. Whitelaw (2008). "A genome-wide screen for modifiers of transgene variegation identifies genes with critical roles in development." Genome Biology **9**(12).

Ban, H., N. Nishishita, N. Fusaki, T. Tabata, K. Saeki, M. Shikamura, N. Takada, M. Inoue, M. Hasegawa, S. Kawamata and S. I. Nishikawa (2011). "Efficient generation of transgene-free human induced pluripotent stem cells (iPSCs) by temperature-sensitive Sendai virus

vectors." Proceedings of the National Academy of Sciences of the United States of America **108**(34): 14234-14239.

Barnett, C. and J. E. Krebs (2011). "WSTF does it all: a multifunctional protein in transcription, repair, and replication." Biochemistry and Cell Biology-Biochimie Et Biologie Cellulaire **89**(1): 12-23.

Barnett, C., O. Yazgan, H. C. Kuo, S. Malakar, T. Thomas, A. Fitzgerald, W. Harbour, J. J. Henry and J. E. Krebs (2012). "Williams Syndrome Transcription Factor is critical for neural crest cell function in Xenopus laevis." Mech Dev **129**(9-12): 324-338.

Barnett, C., O. Yazgan, H. C. Kuo, S. Malakar, T. Thomas, A. Fitzgerald, W. Harbour, J. J. Henry and J. E. Krebs (2012). "Williams Syndrome Transcription Factor is critical for neural crest cell function in Xenopus laevis." Mechanisms of Development **129**(9-12): 324-338.

Bayes, M., L. F. Magano, N. Rivera, R. Flores and L. A. P. Jurado (2003). "Mutational mechanisms of Williams-Beuren syndrome deletions." American Journal of Human Genetics **73**(1): 131-151.

Bellin, M., M. C. Marchetto, F. H. Gage and C. L. Mummery (2012). "Induced pluripotent stem cells: the new patient?" Nat Rev Mol Cell Biol **13**(11): 713-726.

Bellugi, U., L. Lichtenberger, D. Mills, A. Galaburda and J. R. Korenberg (1999). "Bridging cognition, the brain and molecular genetics: evidence from Williams syndrome." Trends in Neurosciences **22**(5): 197-207.

Berg, J. S., N. Brunetti-Pierri, S. U. Peters, S. H. Kang, C. T. Fong, J. Salamone, D. Freedenberg, V. L. Hannig, L. A. Prock, D. T. Miller, P. Raffalli, D. J. Harris, R. P. Erickson, C. Cunniff, G. D. Clark, M. A. Blazo, D. A. Peiffer, K. L. Gunderson, T. Sahoo, A. Patel, J. R. Lupski, A. L. Beaudet and S. W. Cheung (2007). "Speech delay and autism spectrum

behaviors are frequently associated with duplication of the 7q11.23 Williams-Beuren syndrome region." Genet Med **9**(7): 427-441.

Bonilla-Claudio, M., J. Wang, Y. Bai, E. Klysik, J. Selever and J. F. Martin (2012). "Bmp signaling regulates a dose-dependent transcriptional program to control facial skeletal development." Development **139**(4): 709-719.

Brennand, K. J., A. Simone, J. Jou, C. Gelboin-Burkhart, N. Tran, S. Sangar, Y. Li, Y. Mu, G. Chen, D. Yu, S. McCarthy, J. Sebat and F. H. Gage (2011). "Modelling schizophrenia using human induced pluripotent stem cells." Nature **473**(7346): 221-225.

Brennand, K. J., A. Simone, J. Jou, C. Gelboin-Burkhart, N. Tran, S. Sangar, Y. Li, Y. L. Mu, G. Chen, D. Yu, S. McCarthy, J. Sebat and F. H. Gage (2011). "Modelling schizophrenia using human induced pluripotent stem cells." Nature **473**(7346): 221-+.

Cahan, P. and G. Q. Daley (2013). "Origins and implications of pluripotent stem cell variability and heterogeneity." Nat Rev Mol Cell Biol **14**(6): 357-368.

Cahan, P. and G. Q. Daley (2013). "Origins and implications of pluripotent stem cell variability and heterogeneity." Nature Reviews Molecular Cell Biology **14**(6): 357-368.

Campuzano, V., M. Segura-Puimedon, V. Terrado, C. Sanchez-Rodriguez, M. Coustets, M. Menacho-Marquez, J. Nevado, X. R. Bustelo, U. Francke and L. A. Perez-Jurado (2012). "Reduction of NADPH-oxidase activity ameliorates the cardiovascular phenotype in a mouse model of Williams-Beuren Syndrome." PLoS Genet **8**(2): e1002458.

Chahrour, M., S. Y. Jung, C. Shaw, X. Zhou, S. T. Wong, J. Qin and H. Y. Zoghbi (2008). "MeCP2, a key contributor to neurological disease, activates and represses transcription." Science **320**(5880): 1224-1229.

Chambers, S. M., C. A. Fasano, E. P. Papapetrou, M. Tomishima, M. Sadelain and L. Studer (2009). "Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling (vol 27, pg 275, 2009)." Nature Biotechnology **27**(5): 485-485.

Choi, K. D., J. Yu, K. Smuga-Otto, G. Salvagiotto, W. Rehrauer, M. Vodyanik, J. Thomson and I. Slukvin (2009). "Hematopoietic and Endothelial Differentiation of Human Induced Pluripotent Stem Cells." Stem Cells **27**(3): 559-567.

Cobourne, M. T., G. M. Xavier, M. Depew, L. Hagan, J. Sealby, Z. Webster and P. T. Sharpe (2009). "Sonic hedgehog signalling inhibits palatogenesis and arrests tooth development in a mouse model of the nevoid basal cell carcinoma syndrome." Dev Biol **331**(1): 38-49.

Consortium, H. D. i. (2012). "Induced pluripotent stem cells from patients with Huntington's disease show CAG-repeat-expansion-associated phenotypes." Cell Stem Cell **11**(2): 264-278.

Crawford, T. O. and C. A. Pardo (1996). "The neurobiology of childhood spinal muscular atrophy." Neurobiol Dis **3**(2): 97-110.

Davis, R. L., H. Weintraub and A. B. Lassar (1987). "Expression of a single transfected cDNA converts fibroblasts to myoblasts." Cell **51**(6): 987-1000.

Del Campo, M., A. Antonell, L. F. Magano, F. J. Munoz, R. Flores, M. Bayes and L. A. P. Jurado (2006). "Hemizygosity at the NCF1 gene in patients with Williams-Beuren syndrome decreases their risk of hypertension." American Journal of Human Genetics **78**(4): 533-542.

Dennis, J. F., H. Kurosaka, A. Iulianella, J. Pace, N. Thomas, S. Beckham, T. Williams and P. A. Trainor (2012). "Mutations in Hedgehog acyltransferase (Hhat) perturb Hedgehog signaling, resulting in severe acrania-holoprosencephaly-agnathia craniofacial defects." PLoS Genet **8**(10): e1002927.

Dimos, J. T., K. T. Rodolfa, K. K. Niakan, L. M. Weisenthal, H. Mitsumoto, W. Chung, G. F. Croft, G. Saphier, R. Leibel, R. Goland, H. Wichterle, C. E. Henderson and K. Eggan (2008).

"Induced pluripotent stem cells generated from patients with ALS can be differentiated into motor neurons." Science **321**(5893): 1218-1221.

Dobreva, G., M. Chahrour, M. Dautzenberg, L. Chirivella, B. Kanzler, I. Farinas, G. Karsenty and R. Grosschedl (2006). "SATB2 is a multifunctional determinant of craniofacial patterning and osteoblast differentiation." Cell **125**(5): 971-986.

Ebert, A. D., J. Yu, F. F. Rose, Jr., V. B. Mattis, C. L. Lorson, J. A. Thomson and C. N. Svendsen (2009). "Induced pluripotent stem cells from a spinal muscular atrophy patient." Nature **457**(7227): 277-280.

Edelmann, L., A. Prosnitz, S. Pardo, J. Bhatt, N. Cohen, T. Lauriat, L. Ouchanov, P. J. Gonzalez, E. R. Manghi, P. Bondy, M. Esquivel, S. Monge, M. F. Delgado, A. Splendore, U. Francke, B. K. Burton and L. A. McInnes (2007). "An atypical deletion of the Williams-Beuren syndrome interval implicates genes associated with defective visuospatial processing and autism." J Med Genet **44**(2): 136-143.

Enkhmandakh, B., A. V. Makeyev, L. Erdenechimeg, F. H. Ruddle, N. O. Chimge, M. I. Tussie-Luna, A. L. Roy and D. Bayarsaihan (2009). "Essential functions of the Williams-Beuren syndrome-associated TFII-I genes in embryonic development." Proceedings of the National Academy of Sciences of the United States of America **106**(1): 181-186.

Evans, M. J. and M. H. Kaufman (1981). "Establishment in culture of pluripotential cells from mouse embryos." Nature **292**(5819): 154-156.

Ewart, A. K., C. A. Morris, D. Atkinson, W. S. Jin, K. Sternes, P. Spallone, A. D. Stock, M. Leppert and M. T. Keating (1993). "Hemizygosity at the Elastin Locus in a Developmental Disorder, Williams-Syndrome." Nature Genetics **5**(1): 11-16.

Franco, S. J. and U. Muller (2013). "Shaping Our Minds: Stem and Progenitor Cell Diversity in the Mammalian Neocortex." Neuron **77**(1): 19-34.

Friedman, W. F. and W. C. Roberts (1966). "Vitamin D and the supravalvar aortic stenosis syndrome. The transplacental effects of vitamin D on the aorta of the rabbit." Circulation **34**(1): 77-86.

Fusaki, N., H. Ban, A. Nishiyama, K. Saeki and M. Hasegawa (2009). "Efficient induction of transgene-free human pluripotent stem cells using a vector based on Sendai virus, an RNA virus that does not integrate into the host genome." Proceedings of the Japan Academy Series B-Physical and Biological Sciences **85**(8): 348-362.

Ge, X., Y. Ren, O. Bartulos, M. Y. Lee, Z. Yue, K. Y. Kim, W. Li, P. J. Amos, E. C. Bozkulak, A. Iyer, W. Zheng, H. Zhao, K. A. Martin, D. N. Kotton, G. Tellides, I. H. Park, L. Yue and Y. Qyang (2012). "Modeling supravalvular aortic stenosis syndrome with human induced pluripotent stem cells." Circulation **126**(14): 1695-1704.

Gonzalez, F., S. Boue and J. C. I. Belmonte (2011). "Methods for making induced pluripotent stem cells: reprogramming a la carte." Nature Reviews Genetics **12**(4): 231-242.

Gothelf, D., N. Farber, E. Raveh, A. Apter and J. Attias (2006). "Hyperacusis in Williams syndrome: characteristics and associated neuroaudiologic abnormalities." Neurology **66**(3): 390-395.

Grskovic, M., A. Javaherian, B. Strulovici and G. Q. Daley (2011). "Induced pluripotent stem cells--opportunities for disease modelling and drug discovery." Nat Rev Drug Discov **10**(12): 915-929.

Gurdon, J. B. (1962). "The developmental capacity of nuclei taken from intestinal epithelium cells of feeding tadpoles." J Embryol Exp Morphol **10**: 622-640.

Hansen, D. V., J. L. R. Rubenstein and A. R. Kriegstein (2011). "Deriving Excitatory Neurons of the Neocortex from Pluripotent Stem Cells." Neuron **70**(4): 645-660.

Henrichsen, C. N., G. Csardi, M. T. Zabot, C. Fusco, S. Bergmann, G. Merla and A. Reymond (2011). "Using Transcription Modules to Identify Expression Clusters Perturbed in Williams-Beuren Syndrome." Plos Computational Biology **7**(1).

Hinsley, T. A., P. Cunliffe, H. J. Tipney, A. Brass and M. Tassabehji (2004). "Comparison of TFII-I gene family members deleted in Williams-Beuren syndrome." Protein Science **13**(10): 2588-2599.

Hirano, E., R. H. Knutsen, H. Sugitani, C. H. Ciliberto and R. P. Mecham (2007). "Functional rescue of elastin insufficiency in mice by the human elastin gene - Implications for mouse models of human disease." Circulation Research **101**(5): 523-531.

Hirota, H., R. Matsuoka, X. N. Chen, L. S. Salandanan, A. Lincoln, F. E. Rose, M. Sunahara, M. Osawa, U. Bellugi and J. R. Korenberg (2003). "Williams syndrome deficits in visual spatial processing linked to GTF2IRD1 and GTF2I on chromosome 7q11.23." Genet Med **5**(4): 311-321.

Hoogenraad, C. C., B. Koekkoek, A. Akhmanova, H. Krugers, B. Dortland, M. Miedema, A. van Alphen, W. M. Kistler, M. Jaegle, M. Koutsourakis, N. Van Camp, M. Verhoye, A. van der Linden, I. Kaverina, F. Grosveld, C. I. De Zeeuw and N. Galjart (2002). "Targeted mutation of Cyln2 in the Williams syndrome critical region links CLIP-115 haploinsufficiency to neurodevelopmental abnormalities in mice." Nat Genet **32**(1): 116-127.

Iizuka, K. and Y. Horikawa (2008). "ChREBP: A glucose-activated transcription factor involved in the development of metabolic syndrome." Endocrine Journal **55**(4): 617-624.

Inoue, H., N. Nagata, H. Kurokawa and S. Yamanaka (2014). "iPS cells: a game changer for future medicine." EMBO J **33**(5): 409-417.

Jarvinen-Pasley, A., U. Bellugi, J. Reilly, D. L. Mills, A. Galaburda, A. L. Reiss and J. R. Korenberg (2008). "Defining the social phenotype in Williams syndrome: A model for linking gene, the brain, and behavior." Development and Psychopathology **20**(1): 1-35.

Karnik, S. K., B. S. Brooke, A. Bayes-Genis, L. Sorensen, J. D. Wythe, R. S. Schwartz, M. T. Keating and D. Y. Li (2003). "A critical role for elastin signaling in vascular morphogenesis and disease." Development **130**(2): 411-423.

Kaufmann, W. E. and H. W. Moser (2000). "Dendritic anomalies in disorders associated with mental retardation." Cerebral Cortex **10**(10): 981-991.

Kern, J. K., M. H. Trivedi, C. R. Garver, B. D. Grannemann, A. A. Andrews, J. S. Savla, D. G. Johnson, J. A. Mehta and J. L. Schroeder (2006). "The pattern of sensory processing abnormalities in autism." Autism **10**(5): 480-494.

Kruse, K., R. Pankau, A. Gosch and K. Wohlfahrt (1992). "Calcium metabolism in Williams-Beuren syndrome." J Pediatr **121**(6): 902-907.

Kurosaka, H., A. Iulianella, T. Williams and P. A. Trainor (2014). "Disrupting hedgehog and WNT signaling interactions promotes cleft lip pathogenesis." J Clin Invest **124**(4): 1660-1671.

Lam, P. P. L., Y. M. Leung, L. Sheu, J. Ellis, R. G. Tsushima, L. R. Osborne and H. Y. Gaisano (2005). "Transgenic mouse overexpressing syntaxin-1A as a diabetes model." Diabetes **54**(9): 2744-2754.

Lazebnik, M. B., M. I. Tussie-Luna, P. W. Hinds and A. L. Roy (2009). "Williams-Beuren syndrome-associated transcription factor TFII-I regulates osteogenic marker genes." J Biol Chem **284**(52): 36234-36239.

Lee, G., S. M. Chambers, M. J. Tomishima and L. Studer (2010). "Derivation of neural crest cells from human pluripotent stem cells." Nature Protocols **5**(4): 688-701.

Lee, G., E. P. Papapetrou, H. Kim, S. M. Chambers, M. J. Tomishima, C. A. Fasano, Y. M. Ganat, J. Menon, F. Shimizu, A. Viale, V. Tabar, M. Sadelain and L. Studer (2009). "Modelling pathogenesis and treatment of familial dysautonomia using patient-specific iPSCs." Nature **461**(7262): 402-406.

Li, D. Y., B. Brooke, E. C. Davis, R. P. Mecham, L. K. Sorensen, B. B. Boak, E. Eichwald and M. T. Keating (1998). "Elastin is an essential determinant of arterial morphogenesis." Nature **393**(6682): 276-280.

Li, D. Y., G. Faury, D. G. Taylor, E. C. Davis, W. A. Boyle, R. P. Mecham, P. Stenzel, B. Boak and M. T. Keating (1998). "Novel arterial pathology in mice and humans hemizygous for elastin." Journal of Clinical Investigation **102**(10): 1783-1787.

Li, H. H., M. Roy, U. Kuscuoglu, C. M. Spencer, B. Halm, K. C. Harrison, J. H. Bayle, A. Splendore, F. Ding, L. A. Meltzer, E. Wright, R. Paylor, K. Deisseroth and U. Francke (2009). "Induced chromosome deletions cause hypersociability and other features of Williams-Beuren syndrome in mice." Embo Molecular Medicine **1**(1): 50-65.

Liang, G. Y. and Y. Zhang (2013). "Genetic and Epigenetic Variations in iPSCs: Potential Causes and Implications for Application." Cell Stem Cell **13**(2): 149-159.

Ma, H., R. Morey, R. C. O'Neil, Y. He, B. Daughtry, M. D. Schultz, M. Hariharan, J. R. Nery, R. Castanon, K. Sabatini, R. D. Thiagarajan, M. Tachibana, E. Kang, R. Tippner-Hedges, R. Ahmed, N. M. Gutierrez, C. Van Dyken, A. Polat, A. Sugawara, M. Sparman, S. Gokhale, P. Amato, D. P. Wolf, J. R. Ecker, L. C. Laurent and S. Mitalipov (2014). "Abnormalities in human pluripotent cells due to reprogramming mechanisms." Nature **511**(7508): 177-183.

Malenfant, P., X. Liu, M. L. Hudson, Y. Qiao, M. Hrynchak, N. Riendeau, M. J. Hildebrand, I. L. Cohen, A. E. Chudley, C. Forster-Gibson, E. C. Mickelson, E. Rajcan-Separovic, M. E.

Lewis and J. J. Holden (2012). "Association of GTF2i in the Williams-Beuren syndrome critical region with autism spectrum disorders." J Autism Dev Disord **42**(7): 1459-1469.

Marchetto, M. C., C. Carromeu, A. Acab, D. Yu, G. W. Yeo, Y. Mu, G. Chen, F. H. Gage and A. R. Muotri (2010). "A model for neural development and treatment of Rett syndrome using human induced pluripotent stem cells." Cell **143**(4): 527-539.

Marchetto, M. C., B. Winner and F. H. Gage (2010). "Pluripotent stem cells in neurodegenerative and neurodevelopmental diseases." Hum Mol Genet **19**(R1): R71-76.

Martin, G. R. (1981). "Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells." Proc Natl Acad Sci U S A **78**(12): 7634-7638.

Martinelli, M., M. Arlotti, A. Palmieri, L. Scapoli, A. Savoia, M. Di Stazio, F. Pezzetti, E. Masiero and F. Carinci (2008). "Investigation of MYH14 as a candidate gene in cleft lip with or without cleft palate." Eur J Oral Sci **116**(3): 287-290.

Menendez, L., M. J. Kulik, A. T. Page, S. S. Park, J. D. Lauderdale, M. L. Cunningham and S. Dalton (2013). "Directed differentiation of human pluripotent cells to neural crest stem cells." Nat Protoc **8**(1): 203-212.

Menendez, L., M. J. Kulik, A. T. Page, S. S. Park, J. D. Lauderdale, M. L. Cunningham and S. Dalton (2013). "Directed differentiation of human pluripotent cells to neural crest stem cells." Nature Protocols **8**(1): 203-212.

Merla, G., N. Brunetti-Pierri, L. Micale and C. Fusco (2010). "Copy number variants at Williams-Beuren syndrome 7q11.23 region." Hum Genet **128**(1): 3-26.

Merla, G., N. Brunetti-Pierri, L. Micale and C. Fusco (2010). "Copy number variants at Williams-Beuren syndrome 7q11.23 region." Human Genetics **128**(1): 3-26.

Mervis, C. B., J. Dida, E. Lam, N. A. Crawford-Zelli, E. J. Young, D. R. Henderson, T. Onay, C. A. Morris, J. Woodruff-Borden, J. Yeomans and L. R. Osborne (2012). "Duplication of GTF2I results in separation anxiety in mice and humans." Am J Hum Genet **90**(6): 1064-1070.

Mervis, C. B., J. Dida, E. Lam, N. A. Crawford-Zelli, E. J. Young, D. R. Henderson, T. Onay, C. A. Morris, J. Woodruff-Borden, J. Yeomans and L. R. Osborne (2012). "Duplication of GTF2I results in separation anxiety in mice and humans." American journal of human genetics **90**(6): 1064-1070.

Metzis, V., A. D. Courtney, M. C. Kerr, C. Ferguson, M. C. Rondon Galeano, R. G. Parton, B. J. Wainwright and C. Wicking (2013). "Patched1 is required in neural crest cells for the prevention of orofacial clefts." Hum Mol Genet **22**(24): 5026-5035.

Meyer-Lindenberg, A., C. B. Mervis and K. F. Berman (2006). "Neural mechanisms in Williams syndrome: a unique window to genetic influences on cognition and behaviour." Nature Reviews Neuroscience **7**(5): 380-393.

Minoux, M. and F. M. Rijli (2010). "Molecular mechanisms of cranial neural crest cell migration and patterning in craniofacial development." Development **137**(16): 2605-2621.

Miyoshi, N., H. Ishii, H. Nagano, N. Haraguchi, D. L. Dewi, Y. Kano, S. Nishikawa, M. Tanemura, K. Mimori, F. Tanaka, T. Saito, J. Nishimura, I. Takemasa, T. Mizushima, M. Ikeda, H. Yamamoto, M. Sekimoto, Y. Doki and M. Mori (2011). "Reprogramming of Mouse and Human Cells to Pluripotency Using Mature MicroRNAs." Cell Stem Cell **8**(6): 633-638.

Morris, C. A., I. T. Thomas and F. Greenberg (1993). "Williams syndrome: autosomal dominant inheritance." Am J Med Genet **47**(4): 478-481.

Nguyen, H. N., B. Byers, B. Cord, A. Shcheglovitov, J. Byrne, P. Gujar, K. Kee, B. Schule, R. E. Dolmetsch, W. Langston, T. D. Palmer and R. R. Pera (2011). "LRRK2 mutant iPSC-

derived DA neurons demonstrate increased susceptibility to oxidative stress." Cell Stem Cell **8**(3): 267-280.

Nimchinsky, E. A., B. L. Sabatini and K. Svoboda (2002). "Structure and function of dendritic spines." Annual Review of Physiology **64**: 313-353.

O'Leary, J. and L. R. Osborne (2011). "Global analysis of gene expression in the developing brain of Gtf2ird1 knockout mice." PLoS One **6**(8): e23868.

Ohara-Imaizumi, M., T. Fujiwara, Y. Nakamichi, T. Okamura, Y. Akimoto, J. Kawai, S. Matsushima, H. Kawakami, T. Watanabe, K. Akagawa and S. Nagamatsu (2007). "Imaging analysis reveals mechanistic differences between first- and second-phase insulin exocytosis." Journal of Cell Biology **177**(4): 695-705.

Ohno, K., H. Kato, S. Funahashi, T. Hasegawa and K. Sato (2009). "Characterization of CLP36/Elfin/PDLIM1 in the nervous system." J Neurochem **111**(3): 790-800.

Osborne, L. R. (2010). "Animal models of Williams syndrome." Am J Med Genet C Semin Med Genet **154C**(2): 209-219.

Osborne, L. R. (2010). "Animal Models of Williams Syndrome." American Journal of Medical Genetics Part C-Seminars in Medical Genetics **154C**(2): 209-219.

Palmer, S. J., E. S. E. Tay, N. Santucci, T. T. C. Bach, J. Hook, F. A. Lemckert, R. V. Jamieson, P. W. Gunnning and E. C. Hardeman (2007). "Expression of Gtf2ird1, the Williams syndrome-associated gene, during mouse development." Gene Expression Patterns **7**(4): 396-404.

Pankau, R., C. J. Partsch, M. Winter, A. Gosch and A. Wessel (1996). "Incidence and spectrum of renal abnormalities in Williams-Beuren syndrome." Am J Med Genet **63**(1): 301-304.

Park, I. H., N. Arora, H. Huo, N. Maherali, T. Ahfeldt, A. Shimamura, M. W. Lensch, C. Cowan, K. Hochedlinger and G. Q. Daley (2008). "Disease-specific induced pluripotent stem cells." Cell **134**(5): 877-886.

Pasca, S. P., T. Portmann, I. Voineagu, M. Yazawa, A. Shcheglovitov, A. M. Pasca, B. Cord, T. D. Palmer, S. Chikahisa, S. Nishino, J. A. Bernstein, J. Hallmayer, D. H. Geschwind and R. E. Dolmetsch (2011). "Using iPSC-derived neurons to uncover cellular phenotypes associated with Timothy syndrome." Nat Med **17**(12): 1657-1662.

Pasi, C. E., A. Dereli-Oz, S. Negrini, M. Friedli, G. Fragola, A. Lombardo, G. Van Houwe, L. Naldini, S. Casola, G. Testa, D. Trono, P. G. Pelicci and T. D. Halazonetis (2011). "Genomic instability in induced stem cells." Cell Death Differ **18**(5): 745-753.

Phillips, H. M., T. Papoutsi, H. Soenen, P. Ybot-Gonzalez, D. J. Henderson and B. Chaudhry (2012). "Neural crest cell survival is dependent on Rho kinase and is required for development of the mid face in mouse embryos." PLoS One **7**(5): e37685.

Pober, B. R. (2010). "MEDICAL PROGRESS Williams-Beuren Syndrome (vol 362, pg 239, 2010)." New England Journal of Medicine **362**(22): 2142-2142.

Pober, B. R. (2010). "Williams-Beuren syndrome." N Engl J Med **362**(3): 239-252.

Proschel, C., M. J. Blouin, N. J. Gutowski, R. Ludwig and M. Noble (1995). "Limk1 Is Predominantly Expressed in Neural Tissues and Phosphorylates Serine, Threonine and Tyrosine Residues in-Vitro." Oncogene **11**(7): 1271-1281.

Roy, A. L., H. Du, P. D. Gregor, C. D. Novina, E. Martinez and R. G. Roeder (1997). "Cloning of an Inr- and E-box-binding protein, TFII-I, that interacts physically and functionally with USF1." Embo Journal **16**(23): 7091-7104.

Sadler, L. S., L. K. Robinson, K. R. Verdaasdonk and R. Gingell (1993). "The Williams syndrome: evidence for possible autosomal dominant inheritance." Am J Med Genet **47**(4): 468-470.

Saha, K. and R. Jaenisch (2009). "Technical Challenges in Using Human Induced Pluripotent Stem Cells to Model Disease." Cell Stem Cell **5**(6): 584-595.

Sakurai, T., N. P. Dorr, N. Takahashi, L. A. McInnes, G. A. Elder and J. D. Buxbaum (2011). "Haploinsufficiency of Gtf2i, a Gene Deleted in Williams Syndrome, Leads to Increases in Social Interactions." Autism Research **4**(1): 28-39.

Sakurai, T., N. P. Dorr, N. Takahashi, L. A. McInnes, G. A. Elder and J. D. Buxbaum (2011). "Haploinsufficiency of Gtf2i, a gene deleted in Williams Syndrome, leads to increases in social interactions." Autism Res **4**(1): 28-39.

Sanders, S. J., A. G. Ercan-Sencicek, V. Hus, R. Luo, M. T. Murtha, D. Moreno-De-Luca, S. H. Chu, M. P. Moreau, A. R. Gupta, S. A. Thomson, C. E. Mason, K. Bilguvar, P. B. Celestino-Soper, M. Choi, E. L. Crawford, L. Davis, N. R. Wright, R. M. Dhodapkar, M. DiCola, N. M. DiLullo, T. V. Fernandez, V. Fielding-Singh, D. O. Fishman, S. Frahm, R. Garagaloyan, G. S. Goh, S. Kammela, L. Klei, J. K. Lowe, S. C. Lund, A. D. McGrew, K. A. Meyer, W. J. Moffat, J. D. Murdoch, B. J. O'Roak, G. T. Ober, R. S. Pottenger, M. J. Raubeson, Y. Song, Q. Wang, B. L. Yaspan, T. W. Yu, I. R. Yurkiewicz, A. L. Beaudet, R. M. Cantor, M. Curland, D. E. Grice, M. Gunel, R. P. Lifton, S. M. Mane, D. M. Martin, C. A. Shaw, M. Sheldon, J. A. Tischfield, C. A. Walsh, E. M. Morrow, D. H. Ledbetter, E. Fombonne, C. Lord, C. L. Martin, A. I. Brooks, J. S. Sutcliffe, E. H. Cook, Jr., D. Geschwind, K. Roeder, B. Devlin and M. W. State (2011). "Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism." Neuron **70**(5): 863-885.

Schneuwly, S., R. Klemenz and W. J. Gehring (1987). "Redesigning the body plan of Drosophila by ectopic expression of the homoeotic gene Antennapedia." Nature **325**(6107): 816-818.

Schubert, C. and F. Laccone (2006). "Williams-Beuren syndrome: Determination of deletion size using quantitative real-time PCR." International Journal of Molecular Medicine **18**(5): 799-806.

Scott, R. W. and M. F. Olson (2007). "LIM kinases: function, regulation and association with human disease." Journal of Molecular Medicine-Jmm **85**(6): 555-568.

Seibler, P., J. Graziotto, H. Jeong, F. Simunovic, C. Klein and D. Krainc (2011). "Mitochondrial Parkin recruitment is impaired in neurons derived from mutant PINK1 induced pluripotent stem cells." J Neurosci **31**(16): 5970-5976.

Sellner, L. N. and G. R. Taylor (2004). "MLPA and MAPH: New techniques for detection of gene deletions." Human Mutation **23**(5): 413-419.

Shaffer, L. G., G. M. Kennedy, A. S. Spikes and J. R. Lupski (1997). "Diagnosis of CMT1A duplications and HNPP deletions by interphase FISH: Implications for testing in the cytogenetics laboratory." American Journal of Medical Genetics **69**(3): 325-331.

Shi, Y., P. Kirwan and F. J. Livesey (2012). "Directed differentiation of human pluripotent stem cells to cerebral cortex neurons and neural networks." Nat Protoc **7**(10): 1836-1846.

Shi, Y., P. Kirwan, J. Smith, H. P. Robinson and F. J. Livesey (2012). "Human cerebral cortex development from pluripotent stem cells to functional excitatory synapses." Nat Neurosci **15**(3): 477-486, S471.

Singh, S., X. Yin, M. M. Pisano and R. M. Greene (2007). "Molecular profiles of mitogen activated protein kinase signaling pathways in orofacial development." Birth Defects Res A Clin Mol Teratol **79**(1): 35-44.

Slaugenhaupt, S. A., A. Blumenfeld, S. P. Gill, M. Leyne, J. Mull, M. P. Cuajungco, C. B. Liebert, B. Chadwick, M. Idelson, L. Reznik, C. Robbins, I. Makalowska, M. Brownstein, D. Krappmann, C. Scheidereit, C. Maayan, F. B. Axelrod and J. F. Gusella (2001). "Tissue-specific expression of a splicing mutation in the IKBKAP gene causes familial dysautonomia." Am J Hum Genet **68**(3): 598-605.

Smith, A. G., J. K. Heath, D. D. Donaldson, G. G. Wong, J. Moreau, M. Stahl and D. Rogers (1988). "Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides." Nature **336**(6200): 688-690.

Somerville, M. J., C. B. Mervis, E. J. Young, E. J. Seo, M. del Campo, S. Bamforth, E. Peregrine, W. Loo, M. Lilley, L. A. Perez-Jurado, C. A. Morris, S. W. Scherer and L. R. Osborne (2005). "Severe expressive-language delay related to duplication of the Williams-Beuren locus." N Engl J Med **353**(16): 1694-1701.

Sommer, C. A., M. Stadtfeld, G. J. Murphy, K. Hochedlinger, D. N. Kotton and G. Mostoslavsky (2009). "Induced Pluripotent Stem Cell Generation Using a Single Lentiviral Stem Cell Cassette." Stem Cells **27**(3): 543-549.

Sullivan, G. J., D. C. Hay, I. H. Park, J. Fletcher, Z. Hannoun, C. M. Payne, D. Dalgetty, J. R. Black, J. A. Ross, K. Samuel, G. Wang, G. Q. Daley, J. H. Lee, G. M. Church, S. J. Forbes, J. P. Iredale and I. Wilmut (2010). "Generation of Functional Human Hepatic Endoderm from Human Induced Pluripotent Stem Cells." Hepatology **51**(1): 329-335.

Takahashi, K., K. Tanabe, M. Ohnuki, M. Narita, T. Ichisaka, K. Tomoda and S. Yamanaka (2007). "Induction of pluripotent stem cells from adult human fibroblasts by defined factors." Cell **131**(5): 861-872.

Tassabehji, M., P. Hammond, A. Karmiloff-Smith, P. Thompson, S. S. Thorgeirsson, M. E. Durkin, N. C. Popescu, T. Hutton, K. Metcalfe, A. Rucka, H. Stewart, A. P. Read, M.

Maconochie and D. Donnai (2005). "GTF2IRD1 in craniofacial development of humans and mice." Science **310**(5751): 1184-1187.

Thomson, J. A., J. Itskovitz-Eldor, S. S. Shapiro, M. A. Waknitz, J. J. Swiergiel, V. S. Marshall and J. M. Jones (1998). "Embryonic stem cell lines derived from human blastocysts." Science **282**(5391): 1145-1147.

Torniero, C., B. Dalla Bernardina, F. Novara, R. Cerini, C. Bonaglia, T. Pramparo, R. Ciccone, R. Guerrini and O. Zuffardi (2008). "Dysmorphic features, simplified gyral pattern and 7q11.23 duplication reciprocal to the Williams-Beuren deletion." Eur J Hum Genet **16**(8): 880-887.

Van der Aa, N., L. Rooms, G. Vandeweyer, J. van den Ende, E. Reyniers, M. Fichera, C. Romano, B. Delle Chiaie, G. Mortier, B. Menten, A. Destree, I. Maystadt, K. Mannik, A. Kurg, T. Reimand, D. McMullan, C. Oley, L. Brueton, E. M. Bongers, B. W. van Bon, R. Pfund, S. Jacquemont, A. Ferrarini, D. Martinet, C. Schrander-Stumpel, A. P. Stegmann, S. G. Frints, B. B. de Vries, B. Ceulemans and R. F. Kooy (2009). "Fourteen new cases contribute to the characterization of the 7q11.23 microduplication syndrome." Eur J Med Genet **52**(2-3): 94-100.

Velleman, S. L. and C. B. Mervis (2011). "Children with 7q11.23 Duplication Syndrome: Speech, Language, Cognitive, and Behavioral Characteristics and their Implications for Intervention." Perspect Lang Learn Educ **18**(3): 108-116.

Vitale, A. M., N. A. Matigian, S. Ravishankar, B. Bellette, S. A. Wood, E. J. Wolvetang and A. Mackay-Sim (2012). "Variability in the Generation of Induced Pluripotent Stem Cells: Importance for Disease Modeling." Stem Cells Translational Medicine **1**(9): 641-650.

Wang, K. S., X. Liu, Q. Zhang, N. Aragam and Y. Pan (2012). "Parent-of-origin effects of FAS and PDLIM1 in attention-deficit/hyperactivity disorder." J Psychiatry Neurosci **37**(1): 46-52.

Warren, L., P. D. Manos, T. Ahfeldt, Y. H. Loh, H. Li, F. Lau, W. Ebina, P. K. Mandal, Z. D. Smith, A. Meissner, G. Q. Daley, A. S. Brack, J. J. Collins, C. Cowan, T. M. Schlaeger and D. J. Rossi (2010). "Highly Efficient Reprogramming to Pluripotency and Directed Differentiation of Human Cells with Synthetic Modified mRNA." Cell Stem Cell **7**(5): 618-630.

Wilmut, I., A. E. Schnieke, J. McWhir, A. J. Kind and K. H. Campbell (1997). "Viable offspring derived from fetal and adult mammalian cells." Nature **385**(6619): 810-813.

Wong, A. H. and H. H. Van Tol (2003). "Schizophrenia: from phenomenology to neurobiology." Neurosci Biobehav Rev **27**(3): 269-306.

Yamanaka, S. (2012). "Induced pluripotent stem cells: past, present, and future." Cell Stem Cell **10**(6): 678-684.

Yoshimura, K., H. Kitagawa, R. Fujiki, M. Tanabe, S. Takezawa, I. Takada, I. Yamaoka, M. Yonezawa, T. Kondo, Y. Furutani, H. Yagi, S. Yoshinaga, T. Masuda, T. Fukuda, Y. Yamamoto, K. Ebihara, D. Y. Li, R. Matsuoka, J. K. Takeuchi, T. Matsumoto and S. Kato (2009). "Distinct function of 2 chromatin remodeling complexes that share a common subunit, Williams syndrome transcription factor (WSTF) (Retracted article. See vol. 111, pg. 2398, 2014)." Proceedings of the National Academy of Sciences of the United States of America **106**(23): 9280-9285.

Yoshioka, N., E. Gros, H. R. Li, S. Kumar, D. C. Deacon, C. Maron, A. R. Muotri, N. C. Chi, X. D. Fu, B. D. Yu and S. F. Dowdy (2013). "Efficient Generation of Human iPSCs by a Synthetic Self-Replicative RNA." Cell Stem Cell **13**(2): 246-254.

Young, E. J., T. Lipina, E. Tam, A. Mandel, S. J. Clapcote, A. R. Bechard, J. Chambers, H. T. J. Mount, P. J. Fletcher, J. C. Roder and L. R. Osborne (2008). "Reduced fear and aggression and altered serotonin metabolism in Gtf2ird1-targeted mice." Genes Brain and Behavior **7**(2): 224-234.

Yu, J. Y., M. A. Vodyanik, K. Smuga-Otto, J. Antosiewicz-Bourget, J. L. Frane, S. Tian, J. Nie, G. A. Jonsdottir, V. Ruotti, R. Stewart, I. I. Slukvin and J. A. Thomson (2007). "Induced pluripotent stem cell lines derived from human somatic cells." Science **318**(5858): 1917-1920.

Zhao, X., Z. Qu, J. Tickner, J. Xu, K. Dai and X. Zhang (2014). "The role of SATB2 in skeletogenesis and human disease." Cytokine Growth Factor Rev **25**(1): 35-44.