Research Article • DOI: 10.1515/tnsci-2015-0007 • Translational Neuroscience • 6 • 2015 • 59-86

Translational Neuroscience

DE GRUYTER OPEN

ASSOCIATION OF NEW DELETION/DUPLICATION REGION AT CHROMOSOME 1P21 WITH INTELLECTUAL DISABILITY, SEVERE SPEECH DEFICIT AND AUTISM SPECTRUM DISORDER-LIKE BEHAVIOR: AN ALL-IN APPROACH TO SOLVING THE *DPYD* ENIGMA

Lukrecija Brečević[1,2],
Martina Rinčić[1,2,3],
Željka Krsnik[1],
Goran Sedmak[1],
Ahmed B. Hamid[3],
Nadezda Kosyakova[3],
Ivan Galić[4],
Thomas Liehr[3],
Fran Borovečki[2]

[1]*Croatian Institute for Brain Research, University of Zagreb Medical School, Šalata 12, 10000 Zagreb, Croatia*

[2]*Department for Functional Genomics, Center for Translational and Clinical Research, University of Zagreb Medical School, University Hospital Center Zagreb, Šalata 2, 10000 Zagreb, Croatia*

[3]*Jena University Hospital, Friedrich Schiller University, Institute of Human Genetics, Kollegiengasse 10, 07743 Jena, Germany*

[4]*Center for Rehabilitation Stančić, Stančić bb, 10370 Stančić, Croatia*

Received 9 October 2014
accepted 29 December 2014

Abstract

We describe an as yet unreported neocentric small supernumerary marker chromosome (sSMC) derived from chromosome 1p21.3p21.2. It was present in 80% of the lymphocytes in a male patient with intellectual disability, severe speech deficit, mild dysmorphic features, and hyperactivity with elements of autism spectrum disorder (ASD).

Several important neurodevelopmental genes are affected by the 3.56 Mb copy number gain of 1p21.3p21.2, which may be considered reciprocal in gene content to the recently recognized 1p21.3 microdeletion syndrome. Both 1p21.3 deletions and the presented duplication display overlapping symptoms, fitting the same disorder category. Contribution of coding and non-coding genes to the phenotype is discussed in the light of cellular and intercellular homeostasis disequilibrium. In line with this the presented 1p21.3p21.2 copy number gain correlated to 1p21.3 microdeletion syndrome verifies the hypothesis of a cumulative effect of the number of deregulated genes - homeostasis disequilibrium leading to overlapping phenotypes between microdeletion and microduplication syndromes.

Although *miR-137* appears to be the major player in the 1p21.3p21.2 region, deregulation of the *DPYD* (dihydropyrimidine dehydrogenase) gene may potentially affect neighboring genes underlying the overlapping symptoms present in both the copy number loss and copy number gain of 1p21. Namely, the all-in approach revealed that *DPYD* is a complex gene whose expression is epigenetically regulated by long non-coding RNAs (lncRNAs) within the locus. Furthermore, the long interspersed nuclear element-1 (*LINE-1*) *L1MC1* transposon inserted in *DPYD* intronic transcript 1 (*DPYD-IT1*) lncRNA with its parasites, *TcMAR-Tigger5b* and pair of *Alu* repeats appears to be the "weakest link" within the *DPYD* gene liable to break. Identification of the precise mechanism through which *DPYD* is epigenetically regulated, and underlying reasons why exactly the break (*FRA1E*) happens, will consequently pave the way toward preventing severe toxicity to the antineoplastic drug 5-fluorouracil (5-FU) and development of the causative therapy for the dihydropyrimidine dehydrogenase deficiency.

Keywords

• Neuronal homeostasis • Neurodevelopmental genes • Overlapping phenotypes • Common fragile site *FRA1E* • Epigenetics • Non-coding RNAs • Transposons • Tc1/*mariner* family of transposable elements • Human brain transcriptome • Bones and dental anomalies

1. Introduction

We report a case of a neocentric small supernumerary marker chromosome (sSMC) derived from the 1p21.3p21.2 chromosome in order to provide insight into the molecular processes influencing the phenotype. Most of the genes comprised in sSMC(1) are enriched in the developing human brain (*PTBP2*, *DPYD*, *miR-137*, *SNX7*, *LPPR5*, *LOC100129620*, *LPPR4*), as revealed by Yale's genome-wide exon-level transcriptome data base (www.humanbraintranscriptome.org) [1], implicating their role in the processes leading to proper brain organization and functioning.

To date, the genotype-phenotype correlations in genetic disorders have been mostly viewed through coding gene mutations and copy number losses/gains. However, recent studies implicate the important contribution of non-coding RNAs to the phenotype, and a need for an all-in approach in investigating the molecular processes underlying the genetic disorders, necessary for a comprehensive understanding of human disease. Therefore, an all-in approach has been applied in analyzing the genotype-phenotype correlation of 1p21.3p21.2 copy number gain.

2. Material and methods

2.1. Clinical description

The patient was a boy identified by the multiplex ligation-dependent probe amplification (MLPA) screening in individuals with unexplained intellectual disability (ID). He is the fifth child of parents who both function at a borderline intellectual level or mild ID. At the age of 10, the boy and his siblings were placed in foster care homes, as the parents had not been able to take care of them.

According to scant data on his early development, he was born at term after an

* *E-mail: lbr25@hotmail.com; brecevic@hiim.hr*

uneventful pregnancy, and the child's early motor development was unremarkable. His language development, in contrast, was severely delayed. On two occasions in his 2nd and 3rd year of life he was hospitalized for febrile convulsions. At the age of 2, behavioral problems and hyperactivity were recorded, which became more remarkable from the age of 6. At the age of 8 he was diagnosed with a moderate to severe ID, pervasive developmental disorder with elements of autism spectrum disorder (ASD), undeveloped speech, febrile convulsions and nocturnal enuresis.

Physical examination at the age of 11 years and 10 months, revealed a height of 142 cm (17th percentile), a weight of 42 kg (50th percentile) and a head circumference of 52 cm (50th percentile). Presently at the age of 19, his height is 172 cm (25th percentile), weight 60 kg (20th percentile), and head circumference 55 cm (25th percentile). The patient's clinical presentation includes the following features:

Face: The face displays mild dysmorphism with deep-set eyes and hooded eyelids, philtrum with upturned upper lip and an open mouth appearance, narrow high arched palate and maxillary prognathism. In each jaw there are 12 permanent teeth: 4 incisors, 2 canines, 2 premolars (without panoramic tomography it was not possible to determine which premolars are missing, teeth no. 4. or no. 5) and 4 molars (3rd molars/"wisdom teeth" are missing). The first maxillary left incisor is mechanically broken, but both maxillary lateral incisors are undergrown and peg-shaped.

Stature: The shoulders are sloping and dropped. A mild thoracic kyphosis (with a hump resembling a buffalo hump) extends to lumbar lordosis.

Hands and feet: Fingers are long and tapering with bilateral clinodactyly of the 5th finger. Instep is very high and rigid. The first toe is long and widely spaced from the 2nd toe, with partial cutaneous syndactyly of 2nd and 3rd toe. Toes appear pointed. His peculiar gait with a tendency toward toe-walking (most evident when running) resembles cock-walk with equinus gait, during which he leans the body forward.

Genitals exhibit one sided cryptorchidism.

Neurological examination: Psychomotor restlessness and hyperactivity made assessments and evaluations extremely difficult even with multiple examinations. The patient is disoriented, socially unadapted, with very poor concentration and attention. He frequently has an empty gaze, as if he is living in his own world with sudden expressions of a laughing grimace (Fig. 1D and Fig. 1F). Speech is sparse and incomprehensible (at the word level) with echolalia and neologisms. Psychomotor restlessness and hyperactivity, as well as aggressive outbursts, are controlled with neuroleptics. The patient is depicted in Fig. 1A-D at the age of 13 years 10 months, and 19, respectively.

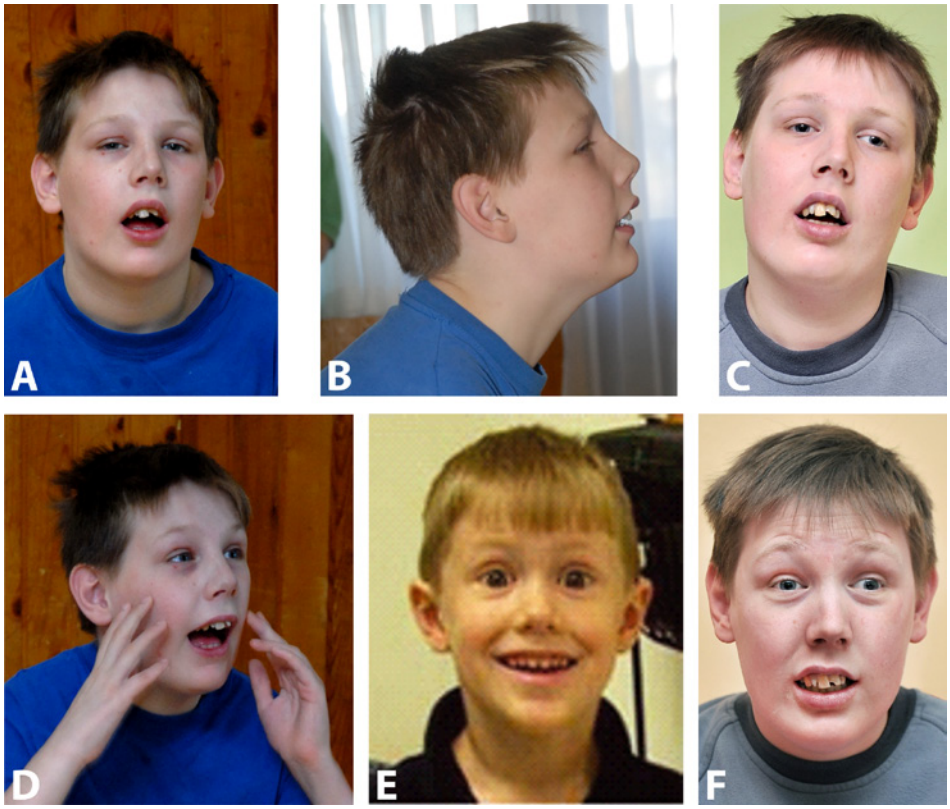


Figure 1. Patient's facial appearance. A-C) The facial appearance (front and lateral view) of the patient at 13/10 years and 19 years, respectively. Note the way the boy is holding his head. D) The facial expression in a sudden laughing/smiling grimace at 13/10 years (F) and at 19 years. Informed consent is obtained from the guardian. E) Note the resemblance of the face expression between the present case and the patient 4 with 10 kb *DPYD* deletion reported by Carter *et al.* [15]. The reuse of the Figure 2f from the paper by Carter *at al.* [15] is kindly provided by © 2010 John Wiley & Sons A/S (license number 3435250114502).

2.2. Cytogenetics and molecular cytogenetics

Banding cytogenetics from peripheral blood lymphocytes was done according to standard procedure. The sSMC was microdissected (glass needle based), the obtained DNA was amplified *in vitro* and labeled by degenerated oligonucleotide-primed polymerase chain reaction (DOP-PCR) as previously reported [2], and applied in standard reverse fluorescence *in situ* hybridization (FISH) [3].

2.3. Multiplex ligation-dependent probe amplification (MLPA)

MLPA probe sets (SALSA MLPA Kits P036-Human Telomere, P070-Human Telomere, P245-Microdeletion Syndromes-1, P297-Microdeletion Syndromes-2, P343-B1 Autism and ME028-PWS/AS) were purchased from MRC-Holland (Amsterdam, The Netherlands) and used according to the manufacturer's protocols. We have used commercially available software, Gene Marker from SoftGenetics (State College, PA, USA) to analyze our data.

2.4. Array-comparative genomic hybridization (arrayCGH)

A genome-targeted copy number profile of the patient's DNA was obtained by subjecting it to microarray analysis using our own custom-designed 8x60 K oligo (60-mer) array platform (Custom CGH Zagreb, ID:061743) manufactured by Agilent Technologies Inc. (Santa Clara, CA, USA), containing most known genes and regions associated with neurodevelopmental disorders (*publication in preparation*). The array was processed following the manufacturer's recommended protocol, and a sex-matched non-disease control sample was used as reference.

3. Results

The present case was singled out during MLPA screening in individuals with unexplained ID, which started with subtelomeres screening (P036 and P070) followed by screening for microdeletion syndromes (P245 and P297) and autism (P343-B1), which all showed normal results. Finally, a ME028-PWS/AS probe set was used to exclude possible Angelman syndrome due to uniparental disomy, and the chromosome culture was set up. The methylation pattern was normal. However, the copy number report in repeated experiments revealed a deletion of *SNRPN* exon1B-b (copy number variation in this region has been described in healthy individuals too), and duplication of the reference 1p21 probe 05330-L04717 (results not shown). Banding cytogenetics disclosed a ring shaped sSMC found in 80% of the lymphocytes (Fig. 2A). Subsequent parallel reverse FISH and arrayCGH, both confirmed what MLPA result already suggested: the sSMC was derived from 1p21 (Fig. 2B and Fig. 2C).

ArrayCGH identified a 3.56 Mb copy number gain on chromosome 1 short arm: arr[hg19] 1p21.3p21.2(96,420,239-99,981,342) x3, containing five annotated RefSeq (http://www.ncbi.nlm.nih.gov/refseq/) coding genes: *PTBP2, DPYD, SNX7, LPPR5* and *LPPR4*, and seven annotated Refseq non-coding RNAs (ncRNAs): antisense long non-coding RNAs (lncRNAs) *DPYD-AS1* and *DPYD-AS2*, microRNAs *miR-137* and *miR-2682*, and uncharacterized long intergenic non-coding RNAs (lincRNAs) *LOC10192824*, *LOC729987* and *LOC100129620* (Fig. 3).

Spatio-temporal expression profiles of the genes involved in 1p21.3p21.2 from the genome-wide exon-level transcriptome data [1] (Fig. 4) revealed their enrichment in human brain during neurogenesis, as well as importance in adult brain functioning (Table 1). In addition, microarray expression data of the transcripts possibly involved in common pathway with the genes comprised in sSMC(1), or commented in the discussion, are shown in Fig. 4.

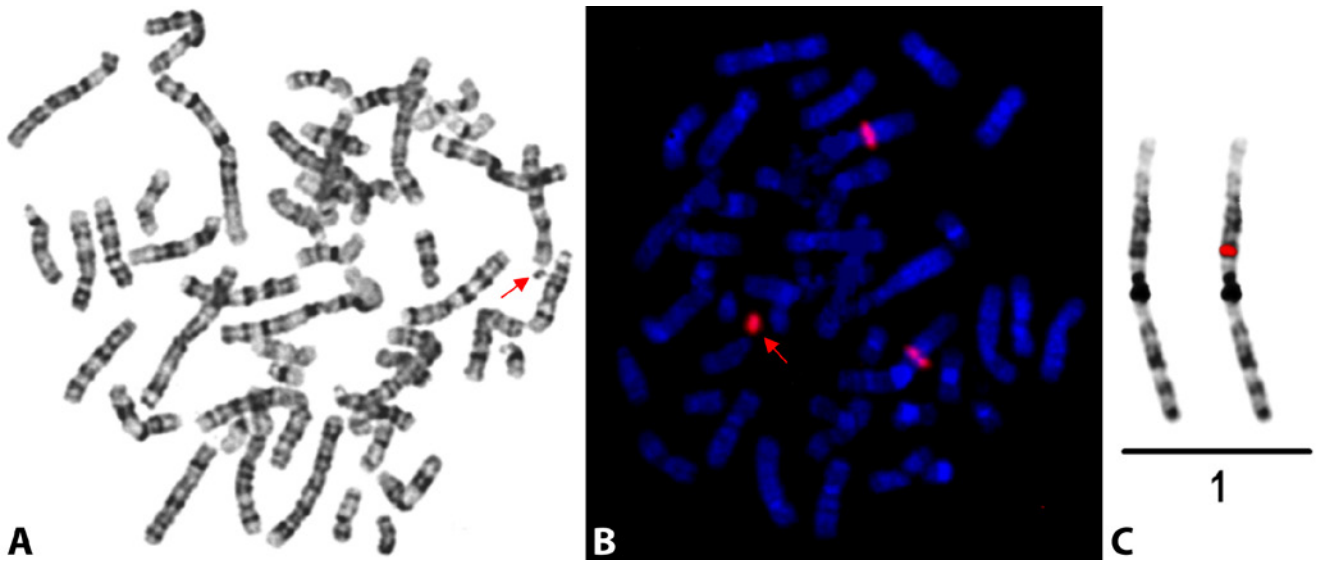


Figure 2. GTG-banded metaphase spread and reverse FISH displaying sSMC(1). A) GTG-banding revealed a karyoytpe 47,XY,+r[80%]/46,XY[20%]; sSMC is labeled by an arrowhead. B) After the microdissection and reverse FISH the red labeled DNA-probe (midi) stained the sSMC itself and a region in the short arm of both chromosomes 1 (arrowhead). C) Inverted DAPI banding shows the mapping of the microdissection derived DNA probe (red; midi) to 1p21.3~21.2.

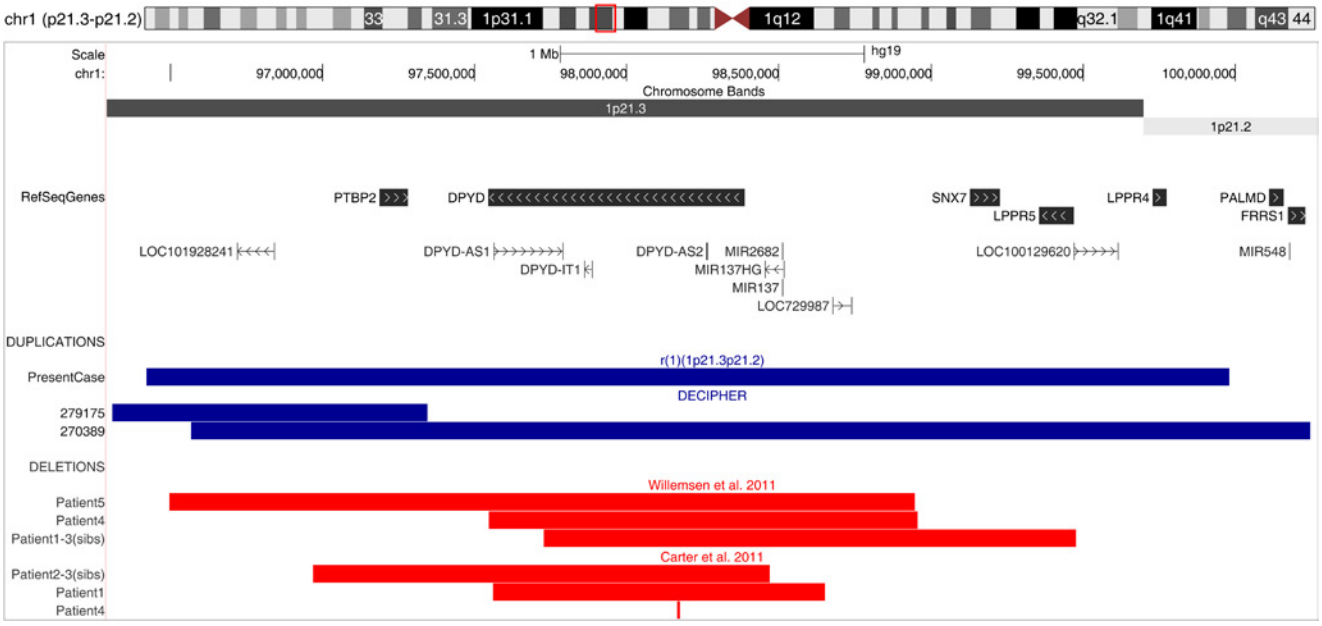


Figure 3. Schematic representations of the deleted and duplicated segments in 1p21. Copy number losses and gains are outlined with the aid of UCSC Genome Browser - Genome Graphs software. Coordinates of the deletions reported by Carter *et al*. [15] and Willemsen *et al*. [16] are converted to hg19. The transcription streams and positions of the coding and non-coding genes along 1p21.3 and adjacent distal 1p21.2 are indicated. The deletions studied by Kuilenburg *et al*. [146] are not included in the figure: patients 1-4 originate from highly consanguineous families; for the patient 5 (~14 Mb, del1p21.3p13.3), no coordinates or most proximal-most distal genes are quoted.

4. Discussion

Small supernumerary marker chromosomes (sSMCs) are found in the general population (0.044%), in infertile (0.125%) and in patients with ID (0.288%) [4]. sSMCs can be present in different shapes, sizes, mosaic states and be derived from different chromosomal regions, predominantly the pericentric ones [5]. Neocentric sSMCs have a centromeric constriction but no detectable alpha-satellite DNA; they "carry newly derived centromeres (or "neocentromeres") that are apparently formed within interstitial chromosomal sites that have not previously been known to express centromere function" [6]. Neocentric sSMCs [7] can be derived from each region of a chromosome. If they come from more distal, i.e. telomeric parts, they often form inverted duplicated shaped sSMCs; in case they are derived from more proximal parts of chromosome arms they are reported as ring shaped sSMCs [8]. Some of them are also formed by a so called McClintock mechanism [9].

We report the eighth case of a neocentric sSMC(1) in clinical practice; additionally, an inverted duplication-shaped sSMC was seen in leukemia as an acquired aberration. Among 8 clinical cases with a neocentric sSMC(1), all except one were ring-shaped like the present case [10]. Three previously reported neocentric sSMC(1) were formed by the McClintock mechanism [9], and one was even derived from a similar region as the present case. However, it was reported as a balanced cytogenetic aberration and no clinical data was available for that case [10, case McCl-01-N-p21/1-1]. Interestingly, the present case is the second one inducing a gain of copy numbers in the short arm of chromosome 1.

In the presented case, sSMC(1) is identified in 80% of lymphocytes, suggesting a possibility of presence in a non-mosaic form and its loss during cell culturing. In addition, it is well known that mosaic cases often display a lower percentage of aberrant cells in lymphocytes than in fibroblasts, the extreme example being Pallister-Killian syndrome [11]. Both fibroblasts and neurons are derived from the same embryonic origin (ectoderm); therefore, it is likely that there are more than 80% of aberrant cells in the brain, even though there is no rule for the distribution of sSMC in different body cells [12]. The expected phenotype with such a high percentage of aberrant cells should not differ much from the non-mosaic form, although the true state of the brain cells remains uncertain. Although it would have been interesting to employ functional magnetic resonance imaging (fMRI) to uncover if any and which cortical areas are affected, due to the risk of performing the procedure under anesthesia, this was not done.

Segmental duplications of 1p are rarely reported. They vary considerably in size and position on 1p and no distinct phenotype has been defined to date for any duplicated segments on 1p. In addition, most previous cases did not have molecular characterization of duplicated segments, so it was not possible to estimate a region of overlap, or genotype-phenotype correlations [reviewed in 13].

4.1. 1p21.3 copy number loss vs. 1p21.3p21.2 copy number gain

Our present case of 1p21.3p21.2 copy number gain is the smallest reported duplication of the proximal short arm of chromosome 1 (Fig. 3). The closest in size are two pathological copy number gains listed in Decipher (http://www.sanger.ac.uk): patient 279175 with 1.04 Mb duplication encompassing

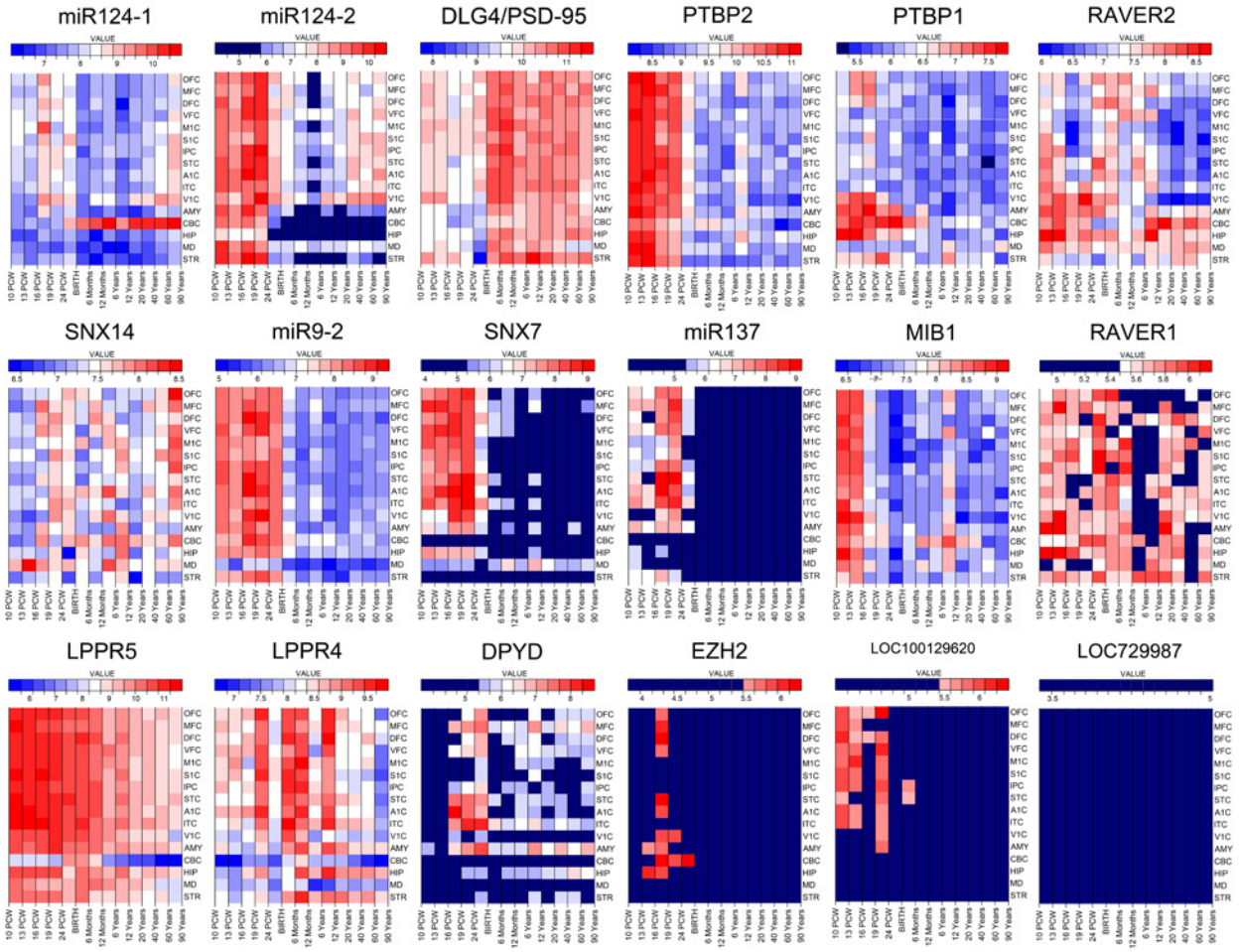


Figure 4. Expression patterns of selected transcripts in the human brain. The microarray expression data were analyzed for affected transcripts in 1p21.3p21.2 and those discussed in the paper in 16 brain regions and 13 developmental periods. Samples colored dark blue are considered unexpressed (cutoff value <5.5). Note that the range of expression intensity is displayed with each transcript profile. Spatiotemporal expression profiles of brain regions and neocortical areas (NCX) were obtained from the Yale's genome-wide exon-level transcriptome data base (www.humanbraintranscriptome.org). MiR124-1 microarray expression data shows overall moderate expression level in the brain, except in cerebral cortex during early mid-fetal period (16-19 PCW). However, from neonatal period until adulthood miR124-1 expression level is consistently high in cerebellum. MiR124-2 expression level is high prenatally during major neurodevelopmental processes in all examined cerebral cortical areas, thalamus, basal ganglia and cerebellum until late mid-fetal period when its expression decreases, especially in cerebellum, hippocampus and striatum. PTBP2 shows similar expression pattern like miR124-2, which is consistent with the finding that miR-124 downregulates PTBP1, leading to upregulation of PTBP2. PTBP2 expression level is consistently high in all brain regions until late mid-fetal period, a crucial neurodevelopmental period for major histogenetic events and formation of neocortical circuits. After 24PCW its expression is downregulated. In contrast, DLG4 (PSP-95) shows moderate expression level in cerebrum and cerebellum until neonatal time, when expression is upregulated in all examined regions (consistent with alternative splicing regulation of PSD-95; PSD-95 is post-transcriptionally repressed by PTBP2, which thus temporarily inhibits the expression of "adult" protein isoforms until neurons have matured) and remain high throughout adulthood. The PTBP1 expression level is consistently low in all neocortical areas, except visual. However, its expression is high prenatally in amygdala, cerebellum and hippocampus until neonatal period when is downregulated. The RAVER1 expression pattern is relatively low throughout lifetime and it shows lowest level during childhood. RAVER2 displays spatiotemporal differential expression pattern in different cortical areas; namely: in prefrontal cortex is highest during perinatal time, while in some cortical areas (like visual) is highly expressed from early fetal until early infancy. RAVER2 expression level is lower from late childhood throughout adulthood in all cortical areas. On a contrary, expression level remains more stable in other brain regions, such as thalamus and basal ganglia. MiR-137 expression is highly expressed prenatally in neocortex and amygdala, but its expression is decreased perinatally and remains low postnatally. According to microarray data, it is not expressed in cerebellum throughout life span. MIB1 shows highest expression during early fetal time in all examined regions of the brain and gradually decreases from early mid-fetal onward in all examined regions, except in cerebellum. SNX7 shows highest expression in neocortical areas, amygdale and hippocampus from early fetal to perinatal period, during the time of major neurodevelopmental processes. Its expression level decreases postnatally. Cerebellum and striatum do not show expression throughout whole life span. MiR9-2 expression level is high from early fetal until neonatal period in all examined brain regions except thalamus. LPPR5 expression is high in all examined brain regions throughout whole life span, except in the cerebellum where it remains lower, except during perinatal period. LPPR4 shows highest expression in neocortical areas, especially during late mid-fetal, neonatal, infancy and childhood during intense synaptogenesis and dendritic differentiation. Its expression is lowest in cerebellum and thalamus throughout life span. DPYD shows substantial spatiotemporal variation in expression pattern throughout lifespan. It is not expressed during early fetal time in the brain and it shows the highest expression level in temporal lobe and amygdale during early and late mid-fetal period. Cerebellum and thalamus do not express it throughout all examined periods. Relatively low LOC100129620 expression can be observed in neocortex from early to late mid-fetal period. LOC729987 is not expressed in any analyzed region or time point. EZH2 is not expressed in the brain throughout lifespan, except in several samples prenatally. SNX14 starts to be expressed at early mid-fetal period and remains to be moderately expressed throughout lifespan. Nomenclature of analyzed brain regions and NCX areas; for the details on ontology see Kang *et al*. [1]: OFC: Orbital prefrontal cortex, DFC: Dorsolateral prefrontal cortex, VFC: Ventrolateral prefrontal cortex, MFC: Medial prefrontal cortex, M1C: Primary motor (M1) cortex, S1C: Primary somatosensory (S1) cortex, IPC: Posterior inferior parietal cortex, A1C: Primary auditory (A1) cortex, STC: Superior temporal cortex, ITC: Inferior temporal cortex, ITC: Inferior temporal cortex, V1C: Primary visual (V1) cortex, HIP: Hippocampus, AMY: Amygdala, STR: Striatum, MD: Mediodorsal nucleus of the thalamus, CBC: Cerebellar cortex.

Table 1. Summary table of the genes involved in 1p21.3p21.2 copy number gain* *For the details and references see the discussion on each gene and Fig. 4.

GENE	PTBP2	DPYD	miR-137	SNX7	LPPR5/PRG5	LPPR4/PRG1
Description	*Polypyrimidine Tract Binding Protein 2*	*Dihydropyrimidine Dehydrogenase*	*MicroRNA 137*	*Sortin Nexin 7*	*Lipid Phosphate Phosphatase-Related Protein Type 5*	*Lipid Phosphate Phosphatase-Related Protein Type 4*
	Multiple transcripts	Multiple transcripts	Highly conserved small noncoding RNA	Multiple transcripts	Multiple transcripts	Multiple transcripts
Tissue specificity	Brain specific (isoform 1 & 2)	Found in most tissues	Neuron-enriched miRNA	Enriched in the brain	Brain & Spinal cord specific	Brain specific
Protein	RNA binding protein	Pyrimidine catabolic enzyme	Non-coding; Binds to multiple target mRNAs	Protein binding	Closely related to LPPR1/PRG3; Mediate LPA activity *in vitro*	Mediate LPA activity in neurons (Hydrolyzes LPA)
Function	Mediates negative regulation of exons splicing	Initial and rate-limiting factor in the pathway of uracil and thymidine catabolism; 5-FU degradation	Translational repression or mRNA degradation; Brain: Silences Mib1 important for neurogenesis	? Exact function (May be involved in several stages of intracellular trafficking)	Involved in neuronal plasticity; Induces filopodia sprouting; Promotes neurite growth; Drives axon elongation	Facilitates axonal outgrowth in the hippocampus; Proper synaptic transmission; Regulator of neuronal plasticity
Expression pattern Human brain transcriptome	Consistent with cross-regulatory network PTBP1-miR124-PTBP2-PSD95	Spatiotemporal variation in expression pattern throughout lifespan	High prenatal expression in neocortex and amygdala	Highest expression in neocortex, amygdale and hippocampus during the time of major neurodevelopmental processes	High in all brain regions throughout whole life span (lower in the cerebellum)	Highest expression in neocortical areas during intense synaptogenesis and dendritic differentiation
Knock out/ Knock down/ Null allele/ Homozygous mutation	Lethal shortly after birth (mice); Neurons in culture-fail to develop	DPD deficiency caused by homozygous or compound heterozygous mutation; Large phenotypic variability; Severe toxicity to 5-FU	Aberrant, enriched dendritic tree of fetal and adult hippocampal neurons	n.a.	Inhibits filopodia formation & neurite growth; Attenuates neurite formation and growth	Juvenile epileptic seizures; Pathological increase of synaptic transmission – Hyperexcitability in CA1 pyramidal neurons
Copy number loss/ Deletion/ Heterozygous mutation/ Haploinsuficiency	PTBP2 at levels half that of WT animals; Target proteins show half reduced-normal expression	Large phenotypic variability; Severe toxicity to 5-FU	Overexpression of validated target proteins	n.a.	n.a.	LPPR4 at levels half that of WT animals; Intermediate increase of excitatory synaptic transmission
Copy number gain/ Duplication/ Overexpression	Repress synaptic activity, spine morphogenesis & reduce PSD-95 transcript *in vitro*	? Exact phenotype effect on brain	Reduces the complexity of dendrites and spine density; 'Deletion effect' of dosage sensitive targets involved in neuronal differentiation	n.a.	Dramatic morphological changes in neuronal cells & non-neuronal cells in vitro	Cognitive deficits observed in mice and men

lincRNA *LOC101928241* and *PTBP2* displaying ASD, muscular hypotonia and strabismus, and patient 270389 with 3.68 Mb duplication which, in addition to all genes comprised in sSMC(1), extends to distal 1p21.2 covering flanking *PALMD*, *FRRS1* and *miR-548* genes, for whom besides the ID no symptoms are commented (Fig. 3). The only case showing phenotypic similarity (primary teeth anomalies, high arched palate, fingers and toes anomalies, peculiar gait, hyperactivity, no speech at 3 years of age; the image of the patient's face also being suggestive of an open mouth appearance and maxillary prognathism) is the case of Utkus *et al*. [14], although the latter duplication appears to be larger.

Unexpectedly, the overlap not only of the neurological and behavioral phenotypes (ID, ASD/ASD-like features and severe to profound speech deficit, febrile/non-febrile seizures), but also of the sum of dysmorphic features seen in differently sized 1p21.3 deletions (deep set eyes, tick lower lip, cleft/high-arched palate, hooded lids) [15, 16], becomes evident when compared to the 1p21.3p21.2 duplication phenotype (Fig. 5).

The 1p21.3p21.2 copy number gain may be considered reciprocal in gene content to the recently recognized 1p21.3 microdeletion syndrome (Fig. 3) [15, 16, www.orpha.net],

characterized by severe speech and language deficit, borderline to moderate and severe ID, ASD features, and minor dysmorphic facial features. Affected individuals have normal gross motor development without major abnormalities. They are often very shy and friendly with a tendency to be overweight.

What could be the rationale behind the fact that both the deletion and duplication of certain genes are capable of producing a similar or overlapping neurological or psychiatric phenotype? According to the hypothesis proposed by Ramocki and Zoghbi [17], a similar or overlapping set of neurological symptoms in reciprocal neurodevelopmental microdeletion and microduplication syndromes can be explained by imbalance of neuronal homeostasis. Briefly, either loss or gain of a certain gene function, which changes synaptic output and neuronal excitability, affects the integrity of the network as a whole, and activates compensation that eventually exhausts homeostatic capacity of the neuronal network and leads to defects of neuronal phenotype and synaptic plasticity. This phenomenon has been described for several genes, such as *MECP2* and *SHANK3*, where loss or gain of function results in overlapping neurological disorders [17].

Therefore, the presented 1p21.3p21.2 copy number gain correlated to 1p21.3 microdeletion syndrome verifies the hypothesis of a cumulative effect of the number of dysregulated genes - homeostasis disequilibrium leading to overlapping phenotypes between microdeletion and microduplication syndromes, since the same conclusion has been drawn from two different points of view.

4.2. Long non-coding RNAs (lncRNAs): LOC10192824, LOC729987 and LOC100129620

The major classes of non-protein coding RNAs (ncRNAs) that are important for the regulation of gene expression include microRNAs (miRNAs), small nucleolar RNAs (snoRNAs), extracellular RNAs (exRNAs), piwi-interacting RNAs (piRNAs), and long-non-coding RNAs (lncRNAs).

Comprehensive analysis of the human transcriptome has revealed that lncRNAs with length >200 nucleotides account for a large fraction of cellular transcripts. The systematization of lncRNAs is still incomplete, as they differ according to genomic localization, size and putative function. Similar to protein coding mRNAs, lncRNA transcripts are capped and polyadenylated, contain multiple exons with large introns and are subject to alternative splicing. Identification of intragenic lncRNAs is hampered due to overlap with protein-coding transcripts or DNA-regulatory elements, and has been originally described as transcription noise [reviewed in 18]. Therefore, they came into the spotlight only recently and their function as key regulators of cellular processes is emerging [19].

Transcription of lncRNAs is cell-type specific and developmentally regulated in the central nervous system (CNS) where they are involved in various roles, such as cell identity, homeostasis, stress responses and synaptic plasticity [20]. Strict temporal and spatial expression of lncRNAs is important for mediating CNS development and function, even though their precise expression pattern and its role are not yet fully known. Many lncRNAs are modulators of gene expression via chromatin modification, and may contain domains for binding other complementary RNAs, protein- and DNA-binding domains that induce conformational changes to other structures in the lncRNA [19].

Hence, it comes as no surprise that some of them are implicated in psychiatric, neurological and neurodegenerative disorders [21]. One of the best studied examples is an antisense long ncRNA *BDNF-AS* [22] that acts as a regulator of expression of brain derived neurotrophic factor (BDNF), important for neuronal growth, maturation and maintenance, whose expression level is reduced in some psychiatric and neurodegenerative disorders, such as Huntington disease [23]. Interestingly, Sauvageau *et al*. [24] reported developmental problems and defects in the cerebral cortex in some of the intergenic lncRNA (lincRNA) knockout mice, thus providing strong evidence of lncRNAs role in brain development.

LOC10192824, LOC100129620 and *LOC729987,* the intergenic lncRNAs that have copy number gain in the present case are yet uncharacterized and their function remains to be elucidated. According to the expression pattern of *LOC100129620* in prenatal cerebral cortex (Fig. 4) we can suggest its role in the regulation of epigenetic dynamics in neurodevelopment.

4.3. MicroRNA miR-137

MicroRNAs constitute a class of small, non-coding RNAs that are involved in a subset of biological processes such as developmental programing, cell proliferation, apoptosis, metabolism, cell differentiation, and morphogenesis [25]. The discovery of microRNAs has led to deeper insights into the regulatory mechanisms of gene expression and their complexity. They function as post-transcriptional regulators of gene expression, primarily through gene silencing by binding to their target mRNAs [26], mediating translational repression or mRNA transcript degradation [27].

A single miRNA typically has multiple, up to several hundred mRNA targets, while a gene can have several target sites for different miRNAs [28-32]. Therefore, miRNAs can control the expression of a number of genes, affecting entire signaling pathways at once leading to a stance that modulation of protein levels by miRNAs represents a key epigenetic regulatory mechanism of gene expression [32-34].

Recent studies suggest that expression of miRNAs and their targets are dynamically regulated, both spatially and temporally, contributing to the diversity and plasticity of our brain [35]. It has been shown that many miRNAs also act locally, at the growth cone or at synapses, modulating synaptic plasticity and neuronal connectivity, thereby contributing to the dynamic spatial organization of axonal and dendritic structures and their function [36, 37].

A neuron-enriched miRNA, *miR-137* plays an important role in the regulation of cell proliferation and differentiation [38-41]. Micro RNA *miR-137* is enriched at the synaptic compartment [36, 37, 39, 42] and regulates neuronal maturation influencing dendritic patterning and spine morphogenesis [43-45] through silencing of Mind bomb one (*Mib1*), an ubiquitin ligase known to be important for neurogenesis [39, 46-48]. There are a large

number of CpG islands in the upstream 2.5 kb promoter region of *miR-137* gene [49], suggesting that its expression is epigenetically regulated.

Overexpression of *miR-137* results in aberrant morphological maturation of neurons by reducing the complexity of dendrites and spine density, both in brain and cultured primary neurons [39]. In contrast, the knock-down of *miR-137* had opposite effects on the dendrite morphogenesis (increases the dendrite length, branch and end points, and number of spines in mouse fetal and adult hippocampal neurons), suggesting that proper expression of *miR-137* is required for the normal morphological differentiation and development of dendrites [39]. Furthermore, all histogenetic processes, including neuronal differentiation, need to be precisely spatially and temporally coordinated in order to establish appropriate synaptic contacts and subsequent proper functioning of cortical neural networks. Hence, resulting either reduced or enriched dendritic tree could lead to misrouted axons and misplaced synaptic contacts that eventually lead to similar or identical abnormal cortical functioning.

A number of developmental and adult brain disorders are associated with abnormal changes in synaptic connectivity and plasticity, [29, 50, 51]. Moreover, a connection has been established between abnormalities in miRNA expression and miRNA-mediated gene regulation and cognitive dysfunction [reviewed in 29 and 35]. In this context, large-scale genome-wide association studies (GWAS) have identified *miR-137* as one of the leading schizophrenia susceptibility genes [52-54].

Consequently, miRNA gene copy number changes due to genomic deletions and duplications are likely to be involved in neurological disorders as well. While studying the clinical effects of chromosome 1p21.3 microdeletions involving *DPYD* and *miR-137*, Willemsen *et al*. [16] found an association with ID and ASD-like behavior. Furthermore, lymphoblastoid cell lines from these patients were found to have reduced levels of *miR-137*. The authors also confirmed that *miR-137* is highly expressed in the hippocampus, occipital cortex, and frontal cortex in human post-mortem tissue, as well as in the synaptosomal fractions of mouse brain preparations, providing further evidence that *miR-137* plays a role in synapse formation during brain development and functioning.

Using the lists of putative and experimentally verified targets, we find no gene in sSMC(1) as the verified target of either *miR-137* or *miR-2682* (http://www.mirbase.org, http://www.targetscan.org, http://mirdb.org). Interestingly, *RAVER2*, ribonucleoprotein PTB-binding 2, whose expression in adult mice is essentially confined to the brain [55, 56] is the highly ranked *miR-137* putative target in miRDB and TargetScan (Agregate PCT 0.96; total context score -0.55). *RAVER1* and *RAVER2* are co-repressors of *PTBP1/PTBP2* (*PTB/nPTB*) homologs, with a modulating function in PTB-mediated RNA processing [57, 58].

Regulation of *RAVER2* by *miR-137* would imply its modulatory role of *PTBP2* expression through downstream mechanism. However, little is still known about the role of *RAVER2* in neurodevelopment due to its restrictive expression pattern and lack of expression in neuronal cell lines [55]. Considering that aberrant expression of miRNAs, leading to either down-regulation or up-regulation of downstream targets, has been implicated in a number of neurodegenerative, neurodevelopmental, as well as psychiatric disorders [59-63], the downstream effect of *miR-137* on the alternative splicing pathway seems probable.

4.4. Polypyrimidine Tract Binding Protein 2 (PTBP2)

PTBP2 protein, encoded by *PTBP2* (*nPTB*) gene is a multifunctional RNA binding protein (shuttling between nucleus and cytoplasm) involved in post-transcriptional regulation of gene expression. PTBP2 shows high tissue-specific expression and shares about 74% amino acid homology [64] with PTBP1 (PTB, HnRNP I), a global repressor of alternative splicing in non-neuronal cells. PTBP1 and PTBP2 display specific non-overlapping expression patterns in the brain; PTBP2 is broadly expressed in the developing mouse brain, including neuronal precursors [65-68], while PTBP1 expression is confined to neuronal precursor cells, glia and other non-neuronal cells [66, 69, 70]. The best known function of the two PTB proteins is the regulation of alternative pre-mRNA splicing patterns, which greatly increase the variety of transcripts indispensable for normal brain development and functioning.

PTBP1 and PTBP2 regulate the synapse formation and maintenance through cross-regulatory network and auto-regulation of expressions [67, 69, 71, 72]. This new genetic regulatory program [65, 73], encompasses three sequential changes in alternative splicing regulation, during which postsynaptic density protein 95 (PSD-95; encoded by *DLG4,* disc large homolog 4 (Drosophila)), essential for synaptic maturation and plasticity, is post-transcriptionally repressed prenatally [67, 68, 71, 72]. The correct switch from general to neuron specific alternative splicing patterns during neuronal differentiation is mediated by neuron specific *miR-124* through down-regulation of *PTBP1* mRNA, which causes a decreased level of PTBP1 protein, and a dramatic increase in PTBP2 protein leading to production of neuron specific protein isoforms [70] (Fig. 4).

Through cross-regulatory network, PTBP2 temporarily inhibits the expression of "adult" protein isoforms until neurons have matured [65, 73, 74], demonstrating an essential role in controlling the brain's early development. These proteins all affect neurite outgrowth, axon guidance, synaptic assembly, and synaptic function; their untimely expression would lead to aberrant neuronal network development [73]. The expression of PTBP2 continues after differentiation, and is present in the brain at moderate levels through adulthood [73] (Fig. 4), but its role in differentiating neurons is not fully understood.

The loss of PTBP2, as demonstrated in *PTBP2* null generated mice (*Ptbp2-/-*), does not greatly affect developmental patterning of CNS. However, post-mitotic neuronal maturation and survival are severely impaired, as a result of misexpression of many protein isoforms affecting neurite growth, synapse formation and synaptic transmission [73]. Similarly, when neurons lacking PTPB2 are grown in culture, they fail to develop correctly and die. Overexpression of PTBP1 and PTBP2 in cultured neurons was shown to repress synaptic activity,

spine formation/morphogenesis, and reduce PSD-95 transcript [68, 73]. The study of Li *et al*. [73] showed that PTBP2 is critical to both embryonic and postnatal brain development.

Constitutional gain of *PTPB2* gene in the presented case and the continuous excess of PTBP2 during neuronal development and throughout the life, certainly to some extent deregulate the cross-regulatory network and specific pattern of expression of both *PTBP1* and *PTBP2*, affecting also the alternative splicing of a number of pre-mRNAs.

Therefore, in the case of *PTBP1* and *PTPB2*, one would expect that loss of one homologue and its protein leads to overexpression of the other homologue resulting in its protein abundance, and vice versa. However, Li *et al*. [73] found no changes in the expression of *PTBP1* in neuronal progenitor cells, astrocytes, ependymal cells, or other non-neuronal cell types in the brain of *PTBP2* null mice (*Ptbp2-/-*). In spite of this finding, it is still possible that even a relatively mild, but long lasting disproportion of two PTB proteins may impair cognitive function as a result of the cumulative effect of multiple disordered gene expression patterns during development.

In fact, the response of each of the multiple targets to constitutional copy number gain/loss of the sequence(s) which regulate their functioning might depend on the sensitivity of the target to the environmental disequilibrium, and the ability of regulatory mechanisms to overcome disordered homeostasis in the cellular and intercellular milieu. This assumption is in line with the finding that mice with *PTBP2* (*Ptbp2-/+*) deletion/copy number loss express *PTBP2* at levels half that of the wild type animals (*Ptbp2+/+*) [73]. The effect of this heterozygous loss of synaptic protein expression and on target transcript splicing was variable. For some targets, protein levels in the heterozygous brains were intermediate between the wild type and homozygous knockout (as one would expect in one gene-one protein relationship). In other cases, the heterozygotes appeared similar to the wild-type mice, expressing close to normal protein levels.

The same holds true for *miR-137*, as both *PTBP2* and *miR-137* affect the expression and function of multiple target sequences in the genome. It has been shown that *miR-137* deletion results in an up-regulation/overexpression of its validated targets [16]. On the other side, bioinformatically predicted *miR-137* targets showed a small but significant down-regulation/lower level of expression of the genes involved in neuronal differentiation [75] following *miR-137* overexpression.

Therefore, the imbalance of sequences with direct and/or downstream influence on multiple genes might, through the cumulative effect of deregulated distant targets (some of which being up-regulated, other being down-regulated, and some being dosage insensitive), be at least partially responsible for the overlapping behavioral and neurologic phenotypes in a number of reciprocal microdeletions and microduplications.

4.5. Bones and dental anomalies: miR-137

Bone organogenesis is a complex process involving the differentiation and crosstalk of multiple cell types, in which the subset of miRNAs has emerged as an important regulator of bone formation and postnatal functions, contributing to every step of osteogenesis [76-78]. The same holds true for tooth development, as the phenotypes associated with mutations in different genes indicate that integrated networks of signaling pathways are the key regulators of tooth morphogenesis [79, 80]. However, the exact mechanism of the regulatory network governed by miRNAs is still poorly understood.

How do the genes within sSMC(1) fit into teeth and osseous abnormalities seen in the present case, and to a lower extent in 1p21.3 deletions? To the best of our knowledge, no known direct connection between the affected genes and the tooth and bone development exists. However, *miR-137* again seems to be a player that affects downstream genes.

Micro RNA *miR-137* is one of the miRNAs that regulate EZH2 (enhancer of zeste homolog 2), a catalytic component of Polycomb repressive complex 2 (PRC2), which epigenetically regulates chromatin structure to silence gene expressions [81, 82]. An increasing body of evidence suggests that EZH2 plays a critical role in stem cell maintenance and differentiation into specific cell lineages, including neurogenesis, adipogenesis and osteogenesis [81, 83, 84]. Recent studies report that craniofacial skeleton formation in higher vertebrates is crucially dependent on epigenetic regulation [85], and that the switch between adipogenesis and osteogenesis can be epigenetically regulated by phosphorylation of EZH2, which suppresses PCR2 catalytic activity [81, 86, 87]. In addition, recent exome-sequencing studies identified missense mutations and in-frame deletions of *EZH2* in patients with Weaver's syndrome, an autosomal dominant disease characterized by learning disabilities, dysmorphic facial features and general overgrowth, which can include tall stature, obesity and macrocephaly [88]. Mutations of *EZH2* are also reported in a cohort of patients with a nonspecific overgrowth syndrome [89]. Interestingly, the patients with 1p21.3 deletions display (borderline) macrocephaly and a tendency to be overweight [15, 16].

Furthermore, among putative *miR-137* targets is the transcription factor Twist-related protein 1 *(Twist-1)* (TargetScan), which together with *Twist-2,* regulates bone formation through transient suppression of *Runx2* gene essential for osteoblastic differentiation and skeletal morphogenesis in mice [90, 91]. *Twist-1* or *Twist-2* deficiency leads to premature osteoblast differentiation [90].

Micro RNA *miR-137* can also influence tooth development. One of the putative targets of *miR-137* is *AXIN1*, which is the key component of canonical Wnt pathway [92]. Vertebrates have two *AXIN* homologous genes (*AXIN1* and *AXIN2*) [93] which appear to be functionally equivalent and interchangeable in Wnt pathway [94]. Loss of Axin2 function is linked to carcinogenesis as well as abnormal bone and tooth development, including hypodontia [94, 95]. In addition to *AXIN1*, three other genes are among putative *miR-137* targets involved in dental development according to TargetScan: *BCOR* (BCL6 Corepressor; transcriptional regulator), *BCORL1* (BCL6 Corepressor-Like 1) [96, 97] and *PVRL1* (*NECTIN1*), respectively [98].

Thus, gain or loss of *miR-137* function could affect both osteogenesis and dentition. However, it appears that osteogenesis and dentition are more affected in 1p21.3/*miR-137*

copy number gain, though the influence of other genes cannot be ruled out.

Micro RNA *miR-137* is an important player in coordinate and complex regulatory events involving a number of genes. In the context of variable sensitivity of target genes/mRNAs/proteins discussed earlier, it is possible to speculate that constitutional gain of *miR-137* function, mimics haploinsufficiency/deletion/copy number loss of dosage sensitive targets leading to disordered homeostasis. Consistent with this is finding of a small but significant down-regulation of *miR-137* targets following *miR-137* over-expression [75]. On the contrary, a copy number loss of *miR-137* can lead to up-regulation of target genes [16], mimicking their overexpression/duplication/copy number gain. More severely affected jaw/bones and teeth in the present case might therefore be due to the "deletion effect" of dosage sensitive target genes, as a result of constitutional gain of *miR-137*.

4.6. Plasticity-related genes LPPR4/PRG1 and LPPR5/PRG5

Lipid phosphate phosphatase-related proteins (LPPRs, LPPR1-5), also referred to as plasticity-related genes (*PRGs*, *PRG1-5*), are a novel class of integral membrane proteins differentially expressed in the developing brain and reexpressed in regenerating axons [99-101], which belong to the lipid phosphate phosphatase (LPP) family. LPPs interfere with lipid phosphate signaling through mediating the extracellular concentration and signal transduction of lipid phosphate esters, lysophosphatidic acid (LPA) and spingosine-1 phosphate (S1P). LPPRs are predominantly expressed in the brain and may act by modifying bioactive lipids and their signaling pathways [101]. However, the exact functional role of LPPRs is still not fully elucidated.

Among several genes that are duplicated in our case report, there are two members of plasticity-related gene family *LPPR4* and *LPPR5* (Fig. 4) important for neuronal outgrowth and synaptic transmission, respectively.

The protein encoded by human lipid phosphate phosphatase-related protein type 4 gene (*LPPR4*), also known as plasticity-related gene 1 (*PRG1*) protein, is specifically expressed in pyramidal neurons, in the membranes of outgrowing axons and dendrites, where it hydrolyzes lysophosphatidic acid / lysophosphatidate (LPA). LPPR4 localizes in hippocampal neurons exclusively at excitatory postsynaptic endings of dendritic spines in rats [102] and mice [103]. During development and regenerative sprouting, LPPR4 attenuates phospholipid-induced axon collapse in outgrowing axons, thereby facilitating axonal outgrowth in the hippocampus. LPPR4 is considered a putative regulator of neuronal plasticity.

It has been shown that lack of LPPR4 leads to juvenile epileptic seizures in mice [103], suggesting LPPR4 dose-dependent pathological increase of synaptic transmission (hyperexcitability) in CA1 pyramidal neurons [103]. Mice with *LPPR4* deletion/copy number loss (*Lppr4*$^{+/-}$) shows LPPR4 expression approximately half that of WT (*Lppr4+/+)* animals, and exhibits intermediate increase of excitatory synaptic transmission (halfway between *Lppr4*$^{-/-}$ and *Lppr4*+/+ mice). Trimbuch *et al*. [103] concluded that the increase in neuronal excitability is due to the lack of LPPR4 at the postsynaptic side. However, seizures can up-regulate *LPPR4* gene and increase LPPR4 protein level in hippocampus and cortex by themselves, suggesting LPPR4 might be detrimental after seizures, contributing to post-seizure cognitive deficits observed in mice and men [100, 104-107].

The importance of LPPR4 for proper synaptic transmission was recently demonstrated by analysis of global gene expression in large groups of patients with refractory mesial temporal lobe epilepsy (RMTLE) [108]. In this study *LPPR4* was identified as one of the hub genes, interacting with the number of genes, in both subgroups of patients (with and without the history of childhood febrile seizures) indicating not only importance for the etiogenesis of the seizures but also for the clinical outcome.

LPPR5/PRG5, novel LPPR with a high homology with *PRG3/LPPR1* is exclusively expressed in nervous system (Fig.4). As shown by Broggini *et al*. [109], LPPR5 induces filopodia formation and axon elongation in primary cortical neurons *in vitro*. Overexpression of *LPPR5* induced morphological changes in both non-neuronal cells and neurons. It has been suggested that LPPR5 is involved in axonal fine-tuning and in the final development of neuronal circuitry.

In summary, even though the exact molecular role of these genes is not fully elucidated, we can speculate, in line with the hypothesis of Ramocki and Zoghbi [17], that the imbalance in their expression level disturbs the fine balance which is necessary for axon fine tuning of neuronal circuits and neural connectivity, thus leading to abnormal neural transmission that could contribute to described neurological phenotype.

4.7. Sortin nexin 7 (SNX7)

Sorting nexin 7 (SNX7) belongs to a large family of proteins involved in intracellular trafficking. Its exact function is unknown and apart from a single study on zebrafish [110], no report on *SNX7* in rodents/humans is cited in PubMed. The mammalian sorting nexin subgroup of 12 genes coding for SNX–BAR proteins (SNX1, SNX2, SNX4–SNX9, SNX18, SNX32 and SNX33) is characterized by two membrane-binding domains: a phosphoinositide-binding Phox homology (PX) domain and a membrane curvature sensing BAR (for Bin–Amphiphysin–Rvs) domain [111-113]. Several SNX–BAR proteins can elicit vesicle-to-tubule transitions *in vitro* and *in vivo*, implicating SNX–BAR proteins as key regulators of tubular-based endosomal sorting [111, 114, 115], an essential process for maintaining cellular homeostasis, with deregulated sorting underlying a variety of pathologies [116, 117].

The contribution of SNX family members to neuronal functioning or disease is poorly understood. However, there are indications that SNXs are disrupted in patients with microcephaly [118], ID [118], and Down syndrome [119]; a link to bipolar disorder [120] and 6q14 microdeletion syndrome [121] has been suggested as well. A very recent finding that sorting nexin14 (*Snx14*) is imprinted in postnatal mouse visual cortical neurons shed new light on imprinting [122]. Imprinted genes can be regulated in specific cell types and developmental stages [123] which make their identification and validation difficult. To overcome the limitation in identifying new

neuron-specific imprinted genes, Huang et al. [122] successfully modified previously employed approaches [124, 125]. SNX14 protein levels increase during mouse brain development exhibiting predominant expression during brain development and maturation; starting in the early mid-fetal period, SNX14 is expressed moderately throughout lifetime in the human brain (Fig. 4). SNX14 localizes to the cytoplasm and dendrites of dissociated mouse cortical neurons where it regulates neuronal intrinsic excitability and promotes synaptic transmission [122].

SNX7 is a putative target of a single broadly conserved microRNA, *miR-9* (TargetScan), one of the most highly expressed microRNAs in the developing and adult vertebrate brain [reviewed in 126]. Functional analyses have revealed *miR-9* as a versatile regulator of neurogenesis, which together with *miR-124*, appears to be the core genetic circuit regulating mitotic exit of neural progenitors and the onset of neuronal differentiation [127]. Recent studies link *miR-9* with a number of neurodegenerative disorders [128]. The spatio-temporal expression pattern of *SNX7* is similar to that of *miR-9-2* throughout the life span; the exception is perinatal downregulation of *SNX7* in cerebellum and striatum suggesting the possibility of being silenced by *miR-9* (Fig. 4).

4.8. Dihydropyrimidine dehydrogenase (DPYD) gene

4.8.1. DPYD: the clinics

The *DPYD* (dihydropyrimidine dehydrogenase) gene encodes an enzyme (DPD), the initial and rate-limiting factor in the pathway of pyrimidine catabolism, also a key enzyme in the degradation of chemotherapeutic drug 5-fluorouracil (5-FU). Mutations in this gene result in a pharmacogenetic disorder, namely dihydropyrimidine dehydrogenase deficiency, showing large phenotypic variability and ranging from no symptoms to a convulsive disorder with motor and mental retardation in homozygous patients. These individuals also have an increased risk to develop potential life-threatening toxicity to 5-fluorouracil (5-FU) [129, reviewed in 130]. However, it still remains unclear how the excess of uracil and thymine relates to the specific neurological problems that affect some of the people with dihydropyrimidine dehydrogenase deficiency.

Although more than 50 mutations have been characterized in *DPYD* gene, the majority of them represent variants with unknown biological and clinical significance [131-133]. A splice-site mutation in intron 14 (c.1905+1G>A, IVS14+1G>A, *DPYD**2A, rs3918290) as the most prevalent [129], together with two nonsynonymous coding variants [130, 133-136], is the only known functional variant significantly associated with 5-FU-related high-grade (III/IV) toxicity, as shown by case-control studies [137, 138]. Recent comprehensive sequencing of the *DPYD*, as well as the haplotype-based analyses, revealed deep intronic variants of *DPYD* gene in patients with severe adverse effects [139, 140]. However, in a significant number of patients with reduced DPD activity, no mutations could be identified in the coding part of *DPYD* [141, 142]. On the other side, the finding that a *DPYD* haplotype free of any mutations was associated with 5-FU toxicity, suggested the presence of additional genetic variations in the noncoding region of *DPYD* [139] and a different underlying mechanism of toxicity. In addition, only 50% of heterozygous carriers of deleterious risk *DPYD* variants develop 5-FU toxicity. Since the reported genetic variants do not account for most DPD deficiency cases, the epigenetic regulation of *DPYD* promoter has been suggested as a potential important mechanism in 5-FU toxicity [143]. However, no firm evidence for *DPYD* promoter hypermethylation has been found so far to corroborate such a premise [144, 145].

Recently, 1p21.3 microdeletion syndrome has been recognized [15, 16], pinpointing *miRNA-137* and/or *DPYD* as underlying causes for the neurological and behavioral phenotype in the affected patients. Here we would like to accentuate the molecular organization of the *DPYD* gene, and indicate the way the gene is regulated.

Is there a way to explain the obvious similarity in the face expression between the patient with 10 kb deletion of *DPYD* gene [15], and our patient with 3.56 Mb duplication of 1p21.3p21.2 (Fig. 1D-F)? There is also a partial overlap of phenotypic features seen in 1p21.3 copy number loss and present 1p21.3p21.2 copy number gain, with phenotypic features seen in the severely affected patient having about 14 Mb deletion [146] (Fig. 5).

One explanation could be that it is just a coincidence. If not, does *miR-137* or *DPYD* deregulation affect the chromatin conformation thus influencing the expression of neighboring genes? Is *miR-137* or *DPYD* responsible for the phenotypic resemblance? The regulation of *DPYD* by *miR-137* seems unlikely, since *DPYD*-001 (NM_000110), the protein coding transcript encompassing full *DPYD* genomic sequence, is not a putative target of any of the broadly conservative miRNAs, and displays no site with higher probability of preferential conservation, thus implicating its functioning is regulated otherwise. In addition, miRNAs preferentially act through distant downstream targets. However, *DPYD*-002 (NM_001160301) could be the target of *miR-137*, as it has an overlapping site with lower probability of preferential conservation, for *miR-137/137ab* and *miR-25/32/92abc/363/363-3p/367*, respectively (TargetScan).

4.8.2. DPYD: molecular organization

In accordance with the GRCh38/hg38 annotation, the main known components of *DPYD* are (Figs. 6-7 and Supplementary Fig. 1):

Protein coding sequence which span over entire reverse (-) 843,317 bp long strand of *DPYD* gene, comprised of 23 exons and processed in four transcripts (protein coding *DPYD*-001 - *DPYD*-003 and retained intron *DPYD*-004).

Two known long non-coding natural antisense RNA genes (NATs), *DPYD-AS1* (227 kb, processed in one transcript *DPYD-AS1*-001 with 5 exons), and *DPYD-AS2* (1.15 kb, processed in two transcripts *DPYD-AS2*-001 with 2 exons, and *DPYD-AS*-002 with 3 exons); both NATs lay on the forward (+) strand of the *DPYD* gene.

The fourth gene is a novel sense intronic lncRNA *DPYD-IT1* gene (*DPYD* intronic transcript1; Gene Symbol: *RP11-359C24.1*; manually annotated by Havana project - Vega 39 Annotations *OTTHUMT*00000095693, *ENSG*00000232878). *DPYD-IT1* span over 26 kb (chr1: 97,394,154-97,420,141 [hg38]/chr1: 97,859,710-97,885,697 [hg19]) within intron 14 (chr1: 97,450,058 97,382,462 [hg38]/chr1: 97,915,614-97,848,018 [hg19]) of reverse (-) *DPYD* strand. *DPYD-IT1* is comprised of two exons and one long intron, and processed in a

401bp long transcript product (*DPYD-IT1*-001).

In addition, in NCBI Annotation Release 106, annotated is a new XR_426733.1/ LOC102723700/NC_018912.2, 7.7 kb long NAT to the forward strand of *DPYD* gene, comprised from 3 exons. In GRCh38 Ensembl genebuild, 9 novel EST protein coding transcripts have been annotated; 6 to the forward strand (ENSESTT00000033931–ENSESTT00000033936), and 3 to the reverse strand (ENSESTT00000033938–ENSESTT00000033940). Interestingly, the 5′-end of 352 kb long ENSESTT00000033940 transcript overlaps with 5′-end of intron 14, and 3′-end overlaps more or less with both, 3′- end of *DPYD*-001 transcript and 5′-end of *DPYD-AS1*.

4.8.3. Natural antisense RNA transcripts (NATs) and intronic lncRNA involved in the regulation of DPYD

Non-coding RNAs involved in the molecular organization of *DPYD* locus undoubtedly indicate a complex and multi-layered regulation of the *DPYD* gene. In addition to the two known NATs, *DPYD-AS1* and *DPYD-AS2*, novel *DPYD-IT1* intronic lncRNA gene and NAT *XR*_426733.1/ LOC102723700/ *NC*_018912.2, are annotated within the *DPYD* locus.

Characterization of complex mechanisms that regulate *DPYD* expression is a valuable effort, since the gene expressions regulated by antisense lncRNAs open the possibilities to reverse the process [147-149], thus offering a completely new approach in treating the disease. Understanding the mechanism by which lncRNAs regulate *DPYD* functioning, will be a step forward in understanding the biological significance of mutations within the gene, which will consequently lead to finding the way to cure the dihydropyrimidine dehydrogenase deficiency and preventing the cytotoxicity of 5-FU.

The following lines are concise notes from the recent work that has been done on NATs and intronic lncRNAs in the regulation of gene functioning.

Natural antisense RNA transcripts (NATs) are lncRNAs which are transcribed from the opposite strand of protein-coding genes [NATs and other lncRNAs are reviewed in 150-153]. It is predicted that up to 70% of protein coding genes in humans are bidirectionally transcribed [154-157]. The primary antisense transcript mRNAs share complementary exons with the related sense transcript, but the degree of complementarity of NATs with corresponding sense transcripts varies greatly [158-160]. Recent studies have shown that antisense RNAs usually regulate complementary sense mRNA by modulating chromatin structure in *cis*, thereby acting as epigenetic regulators of gene expressions and chromatin remodeling.

Many NATs display opposite/reverse expression patterns with their sense transcript counterparts, implying that they carry the potential to induce allele-specific gene silencing [161, 162]. Actually, the occurrence of NATs correlates with genes that show monoallelic expression [163].

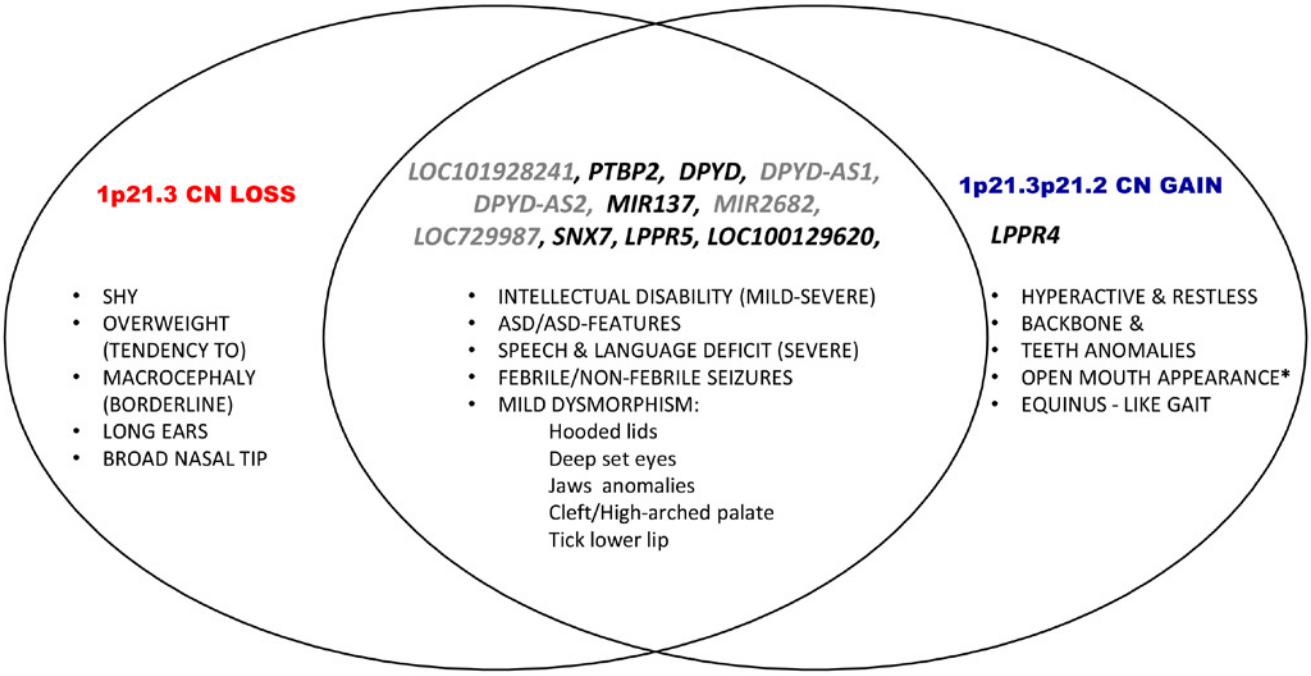


Figure 5. Neurologic and behavioral phenotypes involving physical features in reciprocal 1p21 CN loss/CN gain. Phenotypes unique to CN loss are shown on the left; phenotypes unique to CN gain are shown on the right; phenotypes common to both CN loss and CN gain are shown in the overlapping part of the two ovals. Note that 1p21.3p21.2 copy number gain in addition to overlapping genes with 1p21.3 CN loss encompasses *LPPR4*, the first flanking gene in 1p21.2. *Case 5 from Kuilenburg *et al*. [146] with the deletion involving 1p13.3p21.3, besides profound ID displays macrocephaly, long *prominent/upturned filtrum*, *open mouth appearance*, *tick lower lip*, full nasal tip, *high arched palate* and large lobules. Eruption of his *dentition* was *delayed*, nails were short and thin (in italics are denoted overlapping features with presented 1p21.3p21.2 CN gain; macrocephaly, large lobules, full nasal tip, tick lower lip and high arched palate are the features seen in 1p21.3 CN loss). CN stands for copy number; gray colored are the genes that are either not in the Yale's transcriptome data base (*LOC101928241*, *DPYD-AS1*, *DPYD-AS2*, *miR-2682*), or are not enriched in human brain (*LOC729987*).

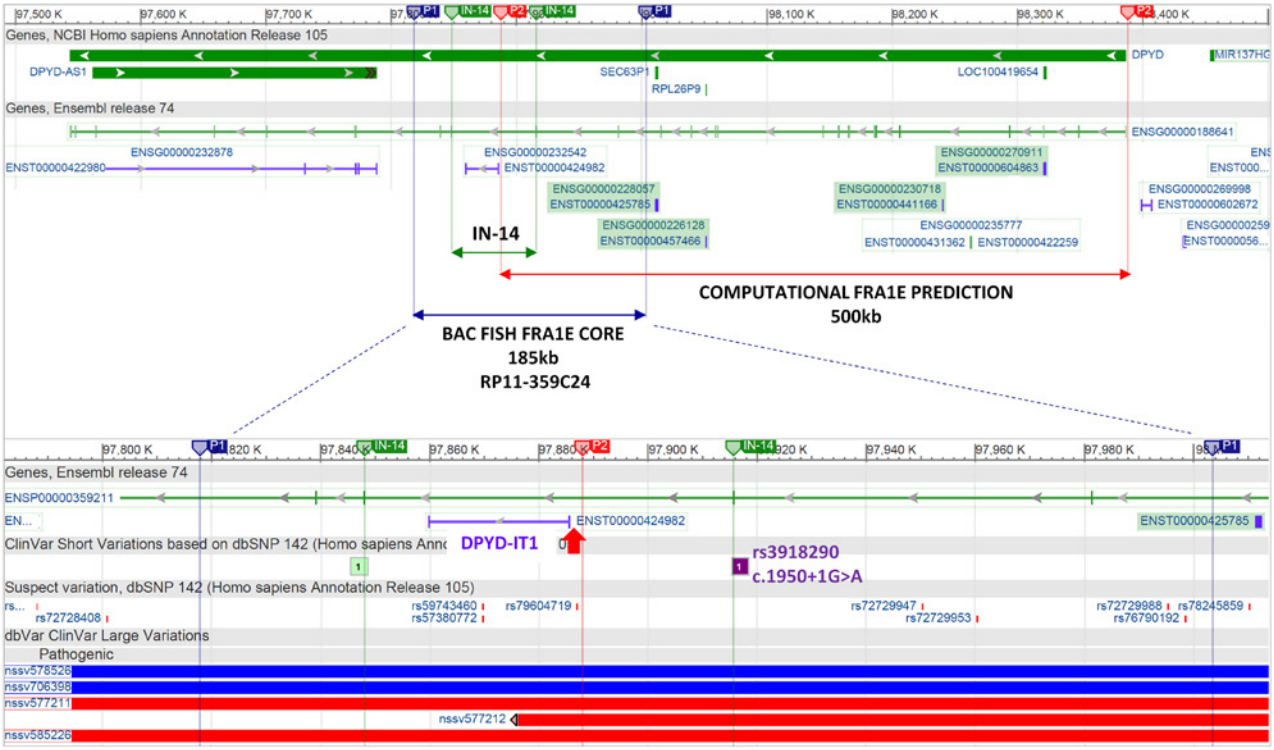


Figure 6. Computational and by BAC FISH predicted coordinates of FRA1E fragile CFS site. FISH predicted *FRA1E* 185 kb core [195] compared with a computational predicted *FRA1E* [196]. Note the overlaps between 5′-end of *DPYD-IT1* and telomeric border of computational *FRA1E* prediction marked by red arrow, and between 5′-end of intron 14 and pathogenic splice variant rs391890/c.1905+1G>A, respectively. Coordinates are in GRCh37/hg19 annotation (The NCBI38 has no more capacity to outline the Ensembl annotated genes). P1 stands for BAC FISH predicted core of *FRA1E*; P2 stands for computational prediction of *FRA1E*.

Recent studies have also shown that NATs work in association with chromatin modifiers, mediating their function through transcriptional and epigenetic regulation, RNA-DNA and RNA-RNA interactions, respectively [150, 151]. The extent of the spread of epigenetic silencing may be related to the CTCF binding factor [164], a multifunctional protein that enables and facilitates higher-order chromatin interactions [165].

An interesting example of transcriptional repression by NATs is INK4b/ARF/INK4a locus regulated by *ANRIL* NAT, where the NAT *ANRIL* participates in the silencing of two very important tumor suppressor genes via two distinct mechanisms. The alteration of these regulatory circuits has been found in different types of cancers [166-168].

NATs are located within many imprinted loci [169-174] and may be directly involved in modulating gene expression within the imprinted cluster. The classic example is the Angelman syndrome gene, *UBE3A*, which is subject to genomic imprinting but not by differential DNA methylation at the promoter region [175]. Instead, *UBE3A* is regulated by its antisense NAT *UBE3A-ATS* in *cis*, which is expressed from the paternally inherited chromosome in the brain and is also imprinted [147, 176].

Although up to 80% of protein coding genes have transcriptionally active introns containing intronic lncRNA genes, little is known about their function [177-179]. Intronic ncRNAs are predominantly associated with the sense strand of the unprocessed mRNA, which is also the case with the *DPYD-IT1*. However, intronic lncRNAs often show expression patterns which are opposite to the processed mRNA [156, 180-182]. This suggests a complex regulatory relationship in which intronic lncRNA transcription is independent from the transcription of protein coding pre-mRNA [177-179]. Intronic lncRNAs may be transcribed from either the sense or antisense strand of the protein-coding gene in which they are encoded [183-185]. Recent work also indicates that many intron-derived RNAs, like many other lncRNAs, function through recruitment of the Polycomb repressive complex 2 (PRC2), leading to subsequent transcriptional repression [186, 187]. Interestingly, *DPYD-IT1* is located within the intron 14, which is the major locus responsible for dihydropyrimidine dehydrogenase deficiency. Namely, the most prominent mutation of the *DPYD* gene that results in severe DPD deficiency is the G to A mutation in the GT 5′-splice recognition site of intron 14 (exon 14-skipping mutation leading to exon 14 deletion). The corresponding mRNA exhibits 165 bp deletion and the enzymatic activity of the translated DPD protein is virtually absent [188].

Another class of lncRNAs, long intergenic ncRNAs (lincRNAs), carries out its regulatory role in *trans*, affecting chromatin conformation and gene expression at distant loci. Transcription

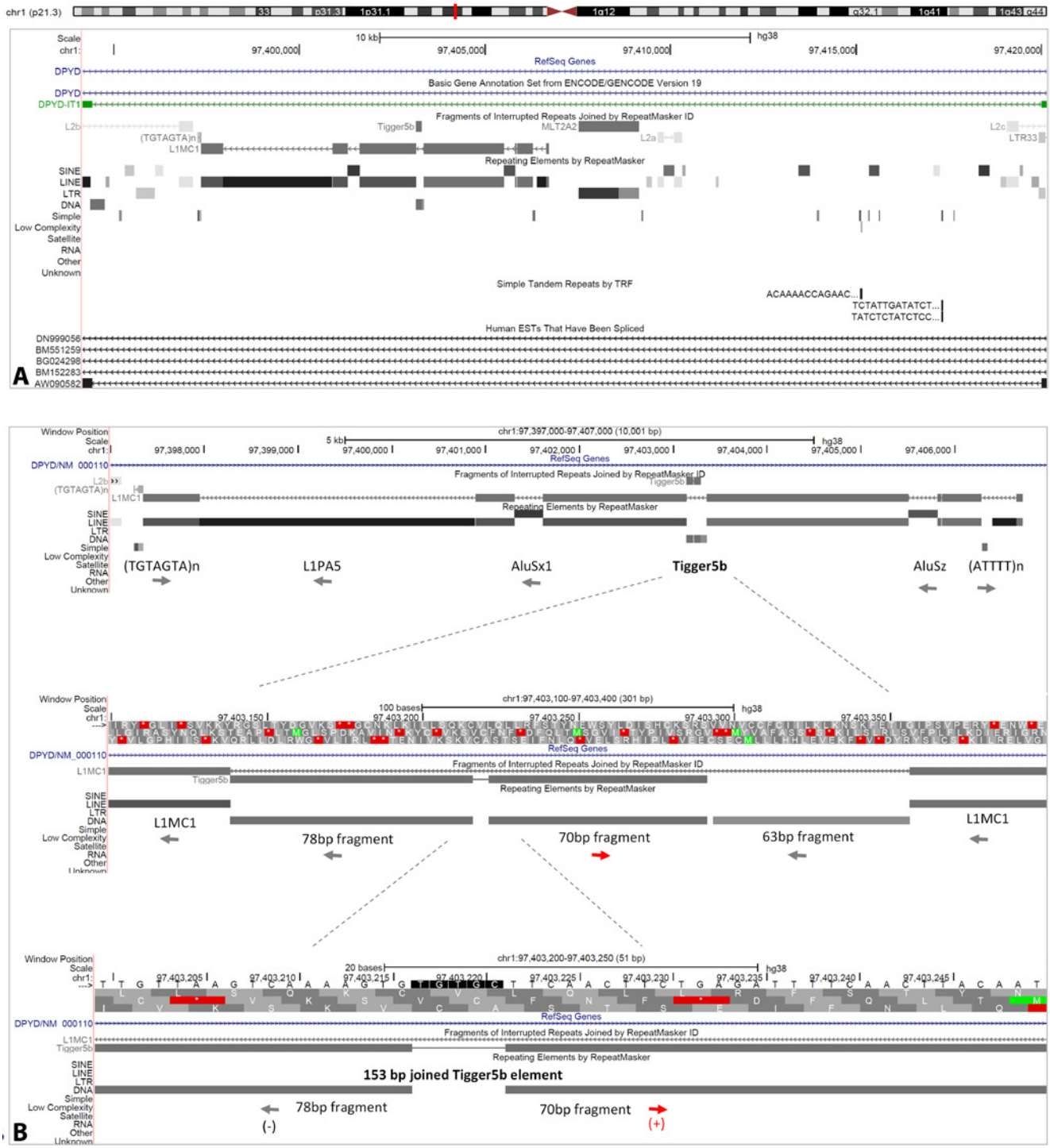


Figure 7. DPYD-IT1 sense intronic lncRNA gene. Overview of repetitive elements comprised in *DPYD-IT1* gene (A). The predicted "weakest link"/center of *FRA1E*: *L1MC1* transposon (chr1: 97,397,345-97,406,722 [hg38]). Note *Tigger5b* DNA element fragments inserted on different genomic strands (B). *GRCh38/hg38 is applied to present the repetitive elements within *DPYD-IT1* since it differs from the older version, GRCh37/hg19.

from such an upstream promoter can negatively or positively affect the expression of a downstream gene. For example, *HOTAIR* is a lincRNA transcribed from the *HOXC* locus that recruits the chromatin remodeling complex, PRC2, to the *HOXD* locus where it creates a repressive chromatin conformation across 40 kb of the locus [189]. Therefore, the epigenetic deregulation of *DPYD* gene may potentially affect neighboring genes underlying the overlapping symptoms present in both the copy number loss and copy number gain of 1p21.

In general, the major role of lncRNAs appears to be the modulation of the epigenetic status of proximal and distal protein-coding genes through *cis*- and *trans*-acting mechanisms regulating chromatin structure over a single gene promoter, a gene cluster, or an entire chromosome [190-194].

All aforementioned possibilities are open, including the one where the *DPYD* and *miR-137* expressions may be mutually affected by one of lincRNAs (*ENST*00000602672.1/RP11-272L13.3, *ENST*00000561881.1/RP11-490G2.2, *MIR-137HG*) near the 5′-end of *DPYD*, acting in *trans*, as all sequences lie on the same, reverse genomic strand and display the same transcriptional direction.

4.8.4. DPYD: common fragile FRA1E site

By partially overlapping bacterial artificial chromosome (BAC) and P1-derived artificial chromosome (PAC) clones, Hormozian *et al*. [195] determined that the *FRA1E* common fragile site extends over 370 kb of genomic region of 1p21 laying within *DPYD* (from intron 8-18). They estimated that the 185 kb region (BAC clone RP11-359C24) of the highest fragility, which accounts for 86% of all observed breaks at *FRA1E*, encompasses the central part of *DPYD*, including exons 13-16.

In a genome-wide analysis of common fragile sites (CFS), Fungtammasan *et al*. [196] computationally predicted the coordinates of 18 CFS including *FRA1E*, which was found to span over 500 kb of the 1p21.3 genomic region. The computationally defined *FRA1E* site was among four out of 18 analyzed CFSs whose coordinates did not overlap with cytogenetically defined coordinates [197, 198]. Discordant results may be explained by both cytogenetic banding and fluorescent mapping methods (on which the multiple standard regression model was derived and estimated), which have inherent technical limitations that contribute to variation among coordinates [196].

In order to correlate both *FRA1E* range predictions, we have converted the computationally revealed coordinates to hg 19 (chr1: 97,887,980-98,387,979) and retrieved the coordinates for the BAC clone RP11-359C24 (chr1: 97,817,963-98,003,396 [hg19]) from UCSC Browser (http://genome.ucsc.edu/). All the data were loaded in Genome Browsers UCSC, Ensembl (www.ensembl.org) and NCBI Variation viewer (www.ncbi.nlm.nih.gov/variation/view/), respectively. Two unexpected findings became apparent, first being the complexity of *DPYD* molecular structure discussed above, and second, the overlap of 5′-end of intronic lncRNA *DPYD-IT1* gene with the telomeric border of computationally predicted *FRA1E* fragile site, that prompted us to continue with the analysis (Fig. 6).

The computer search involving almost the whole 843 kb *DPYD* sequence, both NATs, intron 14 and *DPYD-IT1* (retrieved from NCBI GeneBank and/or Ensembl Browser), for specific DNA repeats and secondary structures that can inhibit replication fork progression [197, 199-202], led us to a conclusion that no definite deduction may be drawn, particularly while current evidence suggests that CFSs are caused by an interplay of multiple genomic factors [203-205].

Nevertheless, questions may be raised whether the *FRA1E* fragile site is a CFS, or where the cluster(s) of loci liable to break precisely map. The most intriguing question is: if such site exists, where the "weakest link"/center of the *FRA1E* maps? It is our belief that the precise characterization of such particular site is as important, as revealing the mechanisms of *DPYD* gene transcriptional regulation; whether *DPYD* is entirely or partly imprinted in the tissue specific manner, or only transiently epigenetically regulated, and in which circumstances.

The immediate focus was on a splice-site, exon skipping mutation at 5′-end of intron 14. Subsequently, other variants also proven to be pathological should be taken into consideration. In such cases only 3-5% [206] of the overall population with true DPD deficiency and additional 2%-3% [140] of the population with partial DPD deficiency due to sequence variation would display the "*FRA1E*" fragile site; these assumptions may be easily verified in case controlled studies. In favor of such assumption speaks the fact that *FRA1E* belongs to the group of CFSs with lower expression than *FRA3B* or *FRA16D* [197], and that many sequence motifs spread throughout the aCFS region may contribute to fragility [204, 207, 208].

Given that the core of the fragile *FRA1E* site entails a 185 kb genomic region (RP11-395C24) between intron 12 and intron 16, the intron 14 is again the most suspected region since the computationally defined telomeric border of *FRA1E* region falls only 2.3 kb upstream from the 5′-end of the *DPYD-IT1* intronic lncRNA gene. The 5′-end of *DPYD-IT1* maps at chr1: 97,885,697 while telomeric border of computationally predicted *FRA1E* maps at chr1: 97,887,980 [hg19] (Fig. 6).

We believe that the site most liable to display a breakage, sort of the "weakest link", maps within *L1MC1* transposable autonomous long interspersed nuclear element-1 (*LINE-1*, *L1*) element, inserted in *DPYD* intronic transcript 1 (*DPYD-IT1*) intronic lncRNA gene (Fig. 7A-B). More precisely, the manner in which the *Tigger5b* DNA element within *L1MC1* is incorporated in the host DNA might be the true "weakest link" within the *DPYD* gene.

LINE-1 transposons

Transposable elements (TEs) are mobile repetitive sequences that make up at least 45% of the human genome [209]. TEs are classified based upon their method of transposition. Class 1 elements transpose via an RNA intermediate through copy-and-paste fashion using reverse transcriptase and include long and short interspersed nuclear elements (LINEs and SINEs), as well as long terminal repeat elements (LTR). Class 2 elements, or DNA transposons, transpose via a DNA intermediate through a cut-and-paste mechanism [210, reviewed in 211].

Although positive contributions of mobile elements to their host genomes are reported, there is growing evidence of the role of TEs in human disease and genetic instability

[reviewed in 212, 213]. The expression of the *L1* retrotransposon can damage the genome through insertional mutagenesis, rearrangements generated by non-allelic homologous recombination (NAHR), and the generation of DNA double-strand breaks (DSBs) [214-220].

In order to show the importance of the L1MC1 retrotransposon, with its parasitic Tigger5b and pair of Alu elements, in the putative regulation of the DPYD gene, the following lines are based (citations) on the papers by Belancio et al. [212], Kines et al. [220], Kines and Belancio [221], Belancio et al. [222], Belancio et al. [223], and Wallace et al. [224]; for comprehensive information see the original papers.

The most active autonomous non-LTR element is long interspersed nuclear element-1, *LINE-1* (*L1*), which contains a number of highly successful parasitic elements [225, 226]. Transcription of a *L1* generates a retrotranspositionally competent, full-length *L1* mRNA [214, 227] and a spectrum of processed *L1*-related RNA products, the majority of which are not capable of retrotransposition [228, 229].

The full-length *L1* mRNA is bicistronic and is influenced by the upstream genomic sequence expression [230]. The functional structure of *L1* element includes promoters, 5′ and 3′ UTRs, two open reading frames (ORF1 and ORF2) required for *L1* retrotransposition [231], and *cis*-acting signals for mRNA processing, with RNA polymerase II (pol II) promoter (sense promoter) located in the beginning of the 5′-untranslated region (UTR) [231, 232]. The antisense *L1* promoter, also present within the 5′UTR, is demonstrated to drive expression of sequences located upstream of the *L1* elements [233, 234]. The biological significance of the antisense promoter is not well established. One of the hypotheses is that its role is to interfere with the transcription initiated within upstream sequences to secure transcription from the sense *L1* promoter. Alternatively, the *L1* antisense promoter is implicated in the production of small interfering RNAs (siRNAs) that inhibit *L1* expression [235, 236]. Both of these promoters can modify the normal gene expression. Independent of the orientation of the *L1* insert (forward or reverse relative to gene expression) they have the potential for "gene breaking" by generating 5′-truncated genomic transcripts [237].

The *L1* promoter activity is regulated by epigenetic modifications [238, 239]. The short- and long-term consequences of *L1* integration (particularly the full-length elements) within or in the vicinity of genes on the epigenetic state and chromatin signature of the gene are not known. Some of the hypotheses dealing with potential contribution of TEs to the epigenetic regulation of the mammalian genome were recently reviewed [240]. *L1* elements have been proposed to potentially influence the selective expression of monoallelically expressed genes due to the enrichment of evolutionarily more recent *LINE-1* elements in the regions surrounding these genes in human and mouse [241]. Furthermore, *L1* promoters contain binding sites for various transcription factors and regulatory proteins that can alter the gene expression in response to various stimuli [242-245]. *L1* sequences can exert their influence on the host gene expression by altering the promoter specificity or strength [246-250].

Furthermore, both sense and antisense *L1* promoters are reported to exhibit tissue-specificity [245, 251]. While no biological significance has been reported to date for the majority of the known *L1*/host gene chimeric mRNAs, cancer-specific *L1*-driven hybrid transcripts were identified in breast and colon cancer cell lines [252].

However, the majority of *L1* loci in the human genome are truncated and incapable of retrotransposition. Although thousands of full-length *L1* loci remain, most are retrotranspositionally-incompetent due to inactivating mutations. However, some of these retrotranspositionally-incompetent *L1* loci previously considered to be inactive and harmless, are indeed expressed [253]. The mutations leading to premature stop codons within the *L1 ORF2* sequence may yield truncated proteins that retain a functional endonuclease domain with the potential to generate low levels of chronic genomic instability by introducing double strand breaks (DSBs) and mobilizing *Alu* sequences [220].

L1 causes insertional mutagenesis through either self retrotransposition or through the mobilization of parasitic non-autonomous transposons, such as *Alu* elements, which rely on the *L1*-encoded ORF2 protein for their propagation [226, 254]. Both *L1*-driven transpositions and *L1*-induced DSBs depend on the endonuclease activity of the *L1* ORF2 protein, which initiates the integration process by nicking the host DNA [219]. Although the origin of the second-strand nick required for completion of the retrotransposition process is unknown, it has been established that expression of the *L1* ORF2 protein containing a functional endonuclease domain results in the formation of DSBs [219, 222, 223, 255]. Importantly, it is estimated that *L1*-induced DSBs are much more frequent than successful *L1*-retrotransposition events [219]. Though the specific consequences of *L1*-induced DSBs are not yet fully known, high mutagenic potential of DSBs in mammalian cells is well documented, contributing to genomic instability and cancer progression [219, reviewed in 256–258].

Ongoing endogenous low-level *L1* activity has been detected in the germ line, as well as in normal human tissues and adult stem cells [259, 260]. Moreover, the *L1* expression is significantly elevated in most human cancers when compared to matched normal tissues [259-264], suggesting a role for *L1* as an endogenous mutagen in somatic tissues.

L1 elements, particularly full-length *L1*s inserted into introns in the *forward orientation* (like *L1MC1* in the intron 14 of *DPYD*), are poorly tolerated [265, 266] and as a result are significantly underrepresented not only within genes, but also in the 5 kb regions flanking human gene boundaries [210, 221, 267].

Insertions of TE within intronic sequences can interfere with normal gene expression through the introduction of functional (i) promoters and their regulatory elements, (ii) polyadenylation (pA) signals, and (iii) splice donor (SD) and acceptor (SA) sites. Besides the effect of TEs on the expression or function of a single gene through direct insertional interference, some TE integration events can also alter gene or cellular pathway function through indirect mechanisms such as regulation of miRNA expression [221].

DNA transposons

The human genome contains about seven major classes of DNA transposons that are virtually all

no longer active, present mostly as fragmented elements in the human genome, and therefore regarded as DNA fossils. Information on human DNA transposons is currently very scarce. However, quite a few functional human genes seem to have originated from DNA transposons, such as genes encoding the *RAG1* and *RAG2* recombinases and the major centromere-binding protein *CENPB*. DNA transposons make up 3% of our genome [210]. Two super families, hAT and Tc1/*mariner*, are predominant in the human DNA transposon population [210, 268, 269].

All complete and autonomous class 2 TEs encode the protein transposases, which are required for insertion and excision. DNA transposons always move on their own, inserting or excising themselves from the genome by means of cut-and-paste mechanism. DNA transposons may also be fully or incompletely inserted into other elements. It has been found that a number of DNA element families, that had copies inserted and integrated into primate-specific *L1* elements, also comprise copies nested into dimeric *Alu* elements, all of which are known to be primate specific [270]; the cohabitation of *L1MC1*, *Tigger5b* and a pair of *Alu* elements comprised in the *DPYD-IT1* gene is an example to this.

DNA transposons have been frequently implicated in chromosomal rearrangements, including deletions, inversions, duplications, translocations, and chromosome breakage mediated by recombination or aberrant transposition events [271-274]. In this context, the recurrent partial *HUWE1* copy number gain which underlies nonsyndromic ID was recently shown to be caused also by NAHR between two adjacent DNA *TcMAR-Tigger2* elements demonstrating that the Xp11.22 region is prone to recombination and replication-based rearrangements [275].

The cut-and-paste transposition mechanism of class II DNA transposons is catalyzed by several transposase enzymes, some of which non-specifically bind to any target site in DNA, whereas others bind to specific DNA sequence targets. The transposase cuts out the DNA transposon (which is then ligated into a new target site) by making a staggered cut at the target site, resulting in single-strand 5′ or 3′ DNA and in most cases in palindromic overhangs or sticky ends. Since the overhangs have to be complementary in order for the ligase to work, the two molecules can only join in one orientation. A transposon or a retroposon that inserts itself into a functional gene will most likely disable that gene, and after a DNA transposon leaves a gene, the resulting gap will probably not be repaired correctly [212].

TEs are also a widely used tool for insertional mutagenesis. Namely, the TE can disrupt the gene's function in a reversible manner, and after the transposase-mediated excision of the DNA transposon, the gene function may be restored. The first synthetic transposon designed for use in vertebrate cells, the *Sleeping Beauty* transposon system, is a Tc1/*mariner*-like transposon. It exists in the human genome as an intron and was activated through reconstruction [276]. The Tc1/*mariner*-class of TEs *Sleeping Beauty* transposon system, the Molecule of the Year 2009 [277] is an example of a transposon system that can be adapted for human gene therapy [278-281] and has been extensively used for identifying cancer genes [282-284].

The *Tigger5b* DNA nested in *L1MC1* element belongs to Tc1/*mariner* superfamily and a family of autonomous DNA transposons and is, like *L1* and *Alu* elements, primate specific [reviewed in 211].

DPYD-IT1: L1MC1, Tigger5b and Alu elements

L1MC1 appears to be a fairly long (9.4 kb) full length *L1* LINE element, nested into the genomic DNA, and could therefore be among about a hundred *L1* transposons still not dormant in the human genome. The status of *L1MC1* activity may depend on the epigenetic (de)regulation of the *DPYD*, but it could also be the part of epigenetic machinery controlling the *DPYD* function since it resides within an intronic lncRNA gene. Though probably inactivated epigenetically or by acquired mutation, the environmental stress could potentially awake *L1MC1* element, since some TEs contain heat-shock like promoters and their rate of transposition increases if the cell is subjected to stress [285].

Moreover, the whole structure of *L1MC1* with nested *Tigger5b* fragments and pair of *Alu* repeats (Fig. 7A-B), offers more than one scenario with detrimental outcome for the host DNA, leading to (irreparable) breaks in case of additional mutations and/or environmental stress. In addition, the organization of *L1MC1* element annotated in GRCh37/hg19 differs from the annotation in GRCh38/hg38, in particular concerning the *Tigger5b* element, suggesting that the whole structure is still awaiting its definite characterization.

The Repeat Masker Track (UCSC Browser, hg38) displays three small fragments of *Tigger5b* DNA transposon: centromeric and telomeric inserted in the reverse/sense strand, while the middle one inserted in the forward/antisense strand, respectively. The fragments are joined in the Interrupted Repeats track in a 153 bp long *Tigger5b* element, composed of two fragments each lying on the different strand (-,+). In between two fragments is a *tgtgc* string of 5 bases (Fig. 7A-B).

The manner in which the *Tigger5b* element is partly inserted in the sense intronic RNA gene, and partly in the antisense DNA strand, suggests an unstable cohabitation between the participants; the situation which potentially may cause problems in the replication processing additionally hampered by agents interfering with DNA transcription, leading to break(s) with sticky ends or overhang which postreplication machinery is unable to repair correctly, ending in the damaged *DPYD* gene. In such circumstances even the whole *Trigger5b* element may be cut out from the genomic sequence with unforeseeable consequences.

The *Tigger5b* is inserted within two *L1MC1* fragments flanked by a full-length *Alu* parasitic TE (lacking ORFs), telomeric *AluSz* (310 bp) and centromeric *AluSx1* (307bp), respectively (Fig. 7A-B). *Alu* elements are the most abundant of all mobile elements in the human genome [210, 286, reviewed in 221] and as such they may act as a template for homologous recombination [287-289]. However, various inherited disorders have been caused by *Alu*-mediated recombination including several types of cancer [287]. Overall, ~0.3% of all human genetic diseases seems to be the result of an *Alu*-mediated unequal homologous recombination. There is also evidence that *Alu* elements inserted into an inverted orientation are more prone to illegitimate recombination [290-293].

Previous discussion on TEs suggests several possibilities how *L1MC1* element with its parasites may potentially affect the functioning of *DPYD*. In addition, the *L1MC1* element seems to be *DPYD* specific. Searching through clinically important aphidicolin-induced CFS associated genes in the RepeatMasker (UCSC Browser), including *PARK2* (MIM 602544)), *DMD* (MIM 300377), *FHIT* (MIM 601153), *WWOX* (MIM 605131), *GRID2* (MIM 602368), *LARGE* (MIM 603590), *CTNNA3* (MIM 607667), and *CNTNAP2* (MIM 604569), no identical *L1* TE element could be found.

5. Conclusion

Even a superficial attempt to comment on the functioning of coding and non-coding sequences comprised in sSMC(1) revealed a complex, and incompletely comprehensible interaction between them, as well as an intricate involvement of these sequences in other cellular and intercellular programs.

Mutation analyses in human disorders have so far been mostly restricted to coding genes. The link between ncRNAs deregulation and neurological diseases is just beginning to emerge. The evolving whole-genome sequencing projects complemented by the arrayCGH studies will undoubtedly reveal that the pathology of neurodevelopmental disorders is indeed affected by mutations and ncRNA copy number gains/losses, thus opening a new avenue for studying the genotype-phenotype correlations in genetic disorders, necessary for a comprehensive understanding of human disease.

Acknowledgments

We would like to thank Goran Šimić for critical assessments and editing the manuscript, and Zdenka Andrijić for the support during the writing process.

Supported in parts by the Croatian Ministry of Science Education and Sport, project No. 108-1081870-1888 (to LB); Business Innovation Croatian Agency – Croatian Institute for Technology BICRO-HIT, project: Targeted aCGH microarrays for diagnostics of neurodevelopmental diseases (to FB); IBRO RHP (to ŽK); and DAAD and Else Kröner Fresenius Stiftung (2011_A42) (to ABH). *Conflict of interest statement*: The authors of this work declare that they have no competing interests.

References

[1] Kang H.J., Kawasawa Y.I., Cheng F., Zhu Y., Xu X., Li M., Sousa A.M., et al., Spatio-temporal transcriptome of the human brain, Nature, 2011, 478, 483-489

[2] Liehr T., Heller A., Starke H., Rubtsov N., Trifonov V., Mrasek K., et al., Microdissection based high resolution multicolor banding for all 24 human chromosomes, Int. J. Mol. Med., 2002, 9, 335-339

[3] Melo J.B., Backx L., Vermeesch J.R., Santos H.G., Sousa A.C., Kosyakova N., et al., Chromosome 5 derived small supernumerary marker: towards a genotype/phenotype correlation of proximal chromosome 5 imbalances, J. App. Genet., 2011, 52, 193-200

[4] Liehr T., Weise A., Frequency of small supernumerary marker chromosomes in prenatal, newborn, developmentally retarded and infertility diagnostics, Int. J. Mol. Med., 2007, 19, 719-731

[5] Liehr T., Claussen U., Starke H., Small supernumerary marker chromosomes (sSMC) in humans, Cytogenet. Genome Res., 2004, 107, 55-67

[6] Choo K.H., Centromere DNA dynamics: latent centromeres and neocentromere formation, Am. J. Hum. Genet., 1997, 61, 1225-1233

[7] Liehr T., Utine G.E., Trautmann U., Rauch A., Kuechler A., Pietrzak J., et al., Neocentric small supernumerary marker chromosomes (sSMC) - three more cases and review of the literature, Cytogenet. Genome Res., 2007, 118, 31-37

[8] Klein E., Rocchi M., Ovens-Raeder A., Kosyakova N., Weise A., Ziegler M., et al., Five novel locations of neocentromeres in human: 18q22.1, Xq27.1~27.2, acro p13, acro p12, and heterochromatin of unknown origin, Cytogenet. Genome Res., 2012, 136, 163-166

[9] Mantzouratou A., Mania A., Apergi M., Laver S., Serhal P., Delhanty J., Meiotic and mitotic behaviour of a ring/deleted chromosome 22 in human embryos determined by preimplantation genetic diagnosis for a maternal carrier, Mol. Cytogenet., 2009, 2, 3

[10] Liehr T., Small supernumerary marker chromosomes, 2014, http://ssmc-tl.com/sSMC.html, accessed December 7, 2015

[11] Yeung A., Francis D., Giouzeppos O., Amor D.J., Pallister-Killian syndrome caused by mosaicism for a supernumerary ring chromosome 12p, Am. J. Med. Genet., 2009, 149A, 505-509

[12] Fickelscher I., Starke H., Schulze E., Ernst G., Kosyakova N., Mkrtchyan H., et al., A further case with a small supernumerary marker chromosome (sSMC) derived from chromosome 1-evidence for high variability in mosaicism in different tissues of sSMC carriers, Prenatal Diag., 2007, 27, 783-785

[13] Piccione M., Antona V., Antona R., Gambino G., Pierluigi M., Malacarne M., et al., Array-CGH defined chromosome 1p duplication in a patient with autism spectrum disorder, mild mental deficiency, and minor dysmorphic features, Am. J. Med. Genet. A., 2010, 152A, 486-489

[14] Utkus A., Sorokina I., Kucinskas V., Röthlisberger B., Balmer D., Brečević L., et al., Duplication of segment 1p21 following paternal insertional translocation, ins(6;1)(q25;p13.3p22.1), J. Med. Genet., 1999, 36, 73-76

[15] Carter M.T., Nikkel S.M., Fernandez B.A., Marshall C.R., Noor A., Lionel A.C., et al., Hemizygous deletions on chromosome 1p21.3 involving the DPYD gene in individuals with autism spectrum disorder, Clin. Genet., 2011, 80, 435-443

[16] Willemsen M.H., Vallès A., Kirkels L.A., Mastebroek M., Olde Loohuis N., Kos A., et al., Chromosome 1p21.3 microdeletions comprising DPYD and MIR137 are associated with intellectual disability, J. Med. Genet., 2011, 48, 810-818

[17] Ramocki M.B., Zoghbi H.Y., Failure of neuronal homeostasis results in common neuropsychiatric phenotypes, Nature, 2008, 455, 912-918

[18] Clark B.S., Blackshaw S., Long non-coding RNA-dependent

transcriptional regulation in neuronal development and disease, Front. Genet., 2014, 5, 164

[19] Mercer T.R., Mattick J.S., Structure and function of long noncoding RNAs in epigenetic regulation, Nat. Struct. Mol. Biol., 2013, 20, 300-307

[20] Cabili M.N., Trapnell C., Goff L., Koziol M., Tazon-Vega B., Regev A., Rinn J.L., Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses, Gene. Dev., 2011, 25, 1915-1927

[21] Vučićević D., Schrewe H., Orom U.A., Molecular mechanisms of long ncRNAs in neurological disorders, Front. Genetics, 2014, 5, 48

[22] Modarresi F., Faghihi M. A., Lopez-Toledano M. A., Fatemi R. P., Magistri M., Brothers S. P., et al., Inhibition of natural antisense transcripts in vivo results in gene-specific transcriptional upregulation, Nat. Biotechnol., 2012, 30, 453-459

[23] Vashishtha M., Ng C. W., Yildirim F., Gipson T. A., Kratter I. H., Bodai L., et al., Targeting H3K4 trimethylation in Huntington disease, Proc. Natl. Acad. Sci. USA, 2013, 110, E3027-E3036

[24] Sauvageau M., Goff L.A., Lodato S., Bonev B., Groff A.F., Gerhardinger C., et al., Multiple knockout mouse models reveal lincRNAs are required for life and brain development, eLIFE, 2013, 2, e01749

[25] Ambros V., The evolution of our thinking about microRNAs, Nat. Med., 2008, 14, 1036-1040

[26] Lewis B.P., Shih I.H., Jones-Rhoades M.W., Bartel D.P., Burge C.B., Prediction of mammalian microRNA targets, Cell, 2003, 115, 787-798

[27] Prosser H.M., Koike-Yusa H., Cooper J.D., Law F.C., Bradley A., 2011. A resource of vectors and ES cells for targeted deletion of microRNAs in mice. Nat. Biotechnol., 29, 840-845.

[28] Lewis B.P., Burge C.B., Bartel D.P., Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets, Cell, 2005, 120, 15-20

[29] Xu B., Hsu P-K., Karayiorgou M., Gogos J.A., MicroRNA dysregulation in neuropsychiatric disorders and cognitive dysfunction, Neurobiol. Dis., 2012, 46 291-301

[30] Yoo A., Staahl B., Chen L., Crabtree G., MicroRNA-mediated switching of chromatin-remodelling complexes in neural development, Nature, 2009, 460, 642-646

[31] Peter M.E., Targeting of mRNAs by multiple miRNAs: the next step, Oncogene, 2010, 29, 2161-2164

[32] Martinez N.J., Gregory R.I., MicroRNA gene regulatory pathways in the establishment and maintenance of ESC identity, Cell Stem Cell, 2010, 7, 31-35

[33] Wu S., Huang S., Ding J., Zhao Y., Liang L., Liu T., Zhan R., He X., Multiple microRNAs modulate p21Cip1/Waf1 expression by directly targeting its 3′ untranslated region, Oncogene, 2010, 29, 2302-2308

[34] Krek A., Grün D., Poy M.N., Wolf R., Rosenberg L., Epstein E.J., et al., Combinatorial microRNA target predictions, Nat. Genet., 2005, 37, 495-500

[35] Olde Loohuis N.F.M, Kos A., Martens G.J.M., Van Bokhoven H., Nadif Kasri N., Aschrafi A., MicroRNA networks direct neuronal development and plasticity, Cell. Mol. Life Sci., 2012, 69, 89-102

[36] Lugli G., Torvik V.I., Larson J., Smalheiser N.R., Expression of microRNAs and their precursors in synaptic fractions of adult mouse forebrain, J. Neurochem., 2008, 106, 650–661

[37] Smalheiser N.R., Lugli G., microRNA regulation of synaptic plasticity, Neuromol. Med., 2009, 11, 133-140

[38] Silber J., Lim D.A., Petritsch C., Persson A.I., Maunakea A.K., Yu M., et al., miR-124 and miR-137 inhibit proliferation of glioblastoma multiforme cells and induce differentiation of brain tumor stem cells, BMC Med., 2008, 6, 14

[39] Smrt R.D., Szulwach K., Pfeiffer R., Li X., Guo W., Pathania M., et al., MicroRNA miR-137 regulates neuronal maturation by targeting ubiquitin ligase mind bomb-1, Stem Cells, 2010, 28, 1060-1070

[40] Sun A.X., Crabtree G.R., Yoo A.S., MicroRNAs: regulators of neuronal fate, Curr. Opin. Cell Biol., 2013, 25, 10

[41] Szulwach K.E., Li X., Smrt R.D., Li Y., Luo Y., Lin L., et al., Cross talk between microRNA and epigenetic regulation in adult neurogenesis, J. Cell Biol., 2010, 189, 127-141

[42] Schratt G., microRNAs at the synapse, Nat. Rev. Neurosci., 2009, 10, 842-849

[43] Lugli G., Torvik V.I., Larson J., Smalheiser N.R. Expression of microRNAs and their precursors in synaptic fractions of adult mouse forebrain, J. Neurochem., 2008, 106, 650-61

[44] Schratt G.M., Tuebing F., Nigh E.A., Kane C.G., Sabatini M.E., Kiebler M., Greenberg M.E., A brain-specific microRNA regulates dendritic spine development, Nature, 2006, 439, 283-289

[45] Siegel G., Obernosterer G., Fiore R., Oehmen M., Bicker S., Christensen M., et al. A functional screen implicates microRNA-138-dependent regulation of the depalmitoylation enzyme APT1 in dendritic spine morphogenesis, Nat. Cell Biol., 2009, 11, 705-716

[46] Choe E.A., Liao L., Zhou J.Y., Cheng D., Duong D.M., Jin P., et al., Neuronal morphogenesis is regulated by the interplay between cyclin-dependent kinase 5 and the ubiquitin ligase mind bomb 1, J. Neurosci., 2007, 27, 9503-9512

[47] Itoh M., Kim C.H., Palardy G., Oda T., Jiang Y.J., Maust D., et al., Mind bomb is a ubiquitin ligase that is essential for efficient activation of Notch signaling by Delta, Dev. Cell, 2003, 4, 67-82

[48] Ossipova O., Ezan J., Sokol S.Y., PAR-1 phosphorylates Mind bomb to promote vertebrate neurogenesis, Dev. Cell, 2009, 17, 222-233

[49] Kunej T., Godnic I., Horvat S., Zorc M., Calin G.A., Cross talk between microRNA and coding cancer genes, Cancer J., 2012, 18, 223-231

[50] Fiala J.C., Spacek J., Harris K.M., Dendritic spine pathology: Cause or consequence of neurological disorders?, Brain Res. Rev., 2002, 39, 29-54

[51] Belmonte M.K., Allen G., Beckel-Mitchener A., Boulanger L.M., Carper R.A., Webb S.J. Autism and abnormal development of brain connectivity, J. Neurosci., 2004, 24, 9228-9231

[52] Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium, Genome-wide association study identifies five new schizophrenia loci, Nat. Genet., 2011, 43, 969-76

[53] Ripke S., O'Dushlaine C., Chambert K., Moran J.L., Kähler A.K., Akterin S., et al., Genome-wide association analysis identifies 13 new risk loci for schizophrenia, Nat. Genet., 2013, 45, 1150-1159

[54] Strazisar M., Cammaerts S., van der Ven K., Forero D.A., Lenaerts A.S., Nordin A., et al., MIR137 variants identified in psychiatric patients affect synaptogenesis and neuronal transmission gene sets, Mol. Psychiatry, 2014, doi: 10.1038/mp.2014.53 [Epub ahead of print]

[55] Kleinhenz B., Fabienke M., Swiniarski S., Wittenmayer N., Kirsch J., Jockusch B.M., et al, Raver2, a new member of the hnRNP family, FEBS Lett., 2005, 579, 4254-4258

[56] Henneberga B., Swiniarskia S., Beckea S., Illenbergera S., A conserved peptide motif in Raver2 mediates its interaction with the polypyrimidine tract-binding protein, Exp. Cell Res., 2010, 316, 966-979

[57] Gromak N., Rideau A.J., Southby A.D., Scadden C., Gooding S., Huttelmaier R.H., et al., The PTB interacting protein raver1 regulates alpha-tropomyosin alternative splicing, EMBO J., 2003, 22, 6356-6364

[58] Romanelli M.R., Diani E., Lievens P.M-J., New Insights into Functional Roles of the Polypyrimidine Tract-Binding Protein, Int. J. Mol. Sci., 2013, 14, 22906-22932

[59] Chang S., Wen S., Chen D., Jin P., Small regulatory RNAs in neurodevelopmental Disorders, Hum. Mol. Genet., 2009, 18, R18-R26

[60] Mitchell K.J., The genetics of neurodevelopmental disease, Curr. Opin. Neurobiol., 2011, 21, 197-203

[61] Henshall D.C., MicroRNA and epilepsy: profiling, functions and potential clinical applications, Curr. Opin. Neurol., 2014, 27, 199-205

[62] Lewis S., Neurological disorders: microRNA gets motoring, Nat. Rev. Neurosci., 2014, 15, 67

[63] Yin J., Lin J., Luo X., Chen Y., Li Z., Ma G., Li K., miR-137: a new player in schizophrenia, Int. J. Mol. Sci., 2014, 15, 3262-71

[64] Oberstrass F.C., Auweter S.D., Erat M., Hargous Y., Henning A., Wenter, P., et al., Structure of PTB bound to RNA: Specific binding and implications for splicing regulation, Science, 2005, 309, 2054-2057

[65] Licatalosi D.D., Yano M., Fak J.J., Mele A., Grabinski S.E., Zhang C., Darnell R.B., Ptbp2 represses adult-specific splicing to regulate the generation of neuronal precursors in the embryonic brain, Gene. Dev., 2012, 15, 26, 1626-1642

[66] Boutz P.L., Stoilov P., Li Q., Lin C.H., Chawla G., Ostrow K., et al., A post-transcriptional regulatory switch in polypyrimidine tract-binding proteins reprograms alternative splicing in developing neurons, Gene. Dev. 2007, 21, 1636-1652

[67] Tang Z.Z., Sharma S., Zheng S., Chawla G., Nikolic J., Black D.L., Regulation of the mutually exclusive exons 8a and 8 in the CaV1.2 calcium channel transcript by polypyrimidine tract-binding protein, J. Biol. Chem., 2011, 286,10007-10016

[68] Zheng S., Gray E.E., Chawla G., Porse B.T., O'Dell T.J., Black D.L., PSD-95 is post-transcriptionally repressed during early neural development by PTBP1 and PTBP2, Nat. Neurosci., 2012, 15, 381-388

[69] Polydorides A.D., Okano H.J., Yang Y.Y., Stefani G., Darnell R.B., A brain-enriched polypyrimidine tract-binding protein antagonizes the ability of Nova to regulate neuron-specific alternative splicing, Proc. Nat. Acad.Sci. USA, 2000, 97, 6350-6355

[70] Makeyev E.V., Zhang J., Carrasco M.A., Maniatis T., The MicroRNA miR-124 promotes neuronal differentiation by triggering brain-specific alternative pre-mRNA splicing, Mol. Cell, 2007, 27, 435-448

[71] Markovtsov V., Nikolic J.M., Goldman J.A., Turck C.W., Chou M.Y., Black D.L., Cooperative assembly of an hnRNP complex induced by a tissue-specific homolog of polypyrimidine tract binding protein, Mol. Cell. Biol., 2000, 20, 7463-7479

[72] Keppetipola N., Sharma S., Li Q., Black D.L., Neuronal regulation of pre-mRNA splicing by polypyrimidine tract binding proteins, PTBP1 and PTBP2, Crit. Rev. Biochem. Mol., 2012, 47, 360-378

[73] Li Q., Zheng S., Han A., Lin C-H., Stoilov P., Fu X-D., et al.,The splicing regulator PTBP2 controls a program of embryonic splicing required for neuronal maturation, eLIFE, 2014, 3, 01201

[74] Wollerton M.C., Gooding C., Wagner E.J., Garcia-Blanco M.A., Smith C.W., Autoregulation of polypyrimidine tract binding protein by alternative splicing leading to nonsense-mediated decay, Mol. Cell, 2004, 13, 91-100

[75] Hill M.J., Donocik J.D., NuamahR.A., Mein C.A., Sainz-Fuertes R., Bray N.J., Transcriptional consequences of schizophrenia candidate miR-137 manipulation in human neural progenitor cells, Schizophr. Res., 2014, 53, 225-230

[76] Dong S., Yang B., Guo H., Kang F., MicroRNAs regulate osteogenesis and chondrogenesis, Biochem. Bioph. Res. Commun., 2012, 418, 587-591

[77] Lian J.B., Stein G.S., van Wijnen A.J., Stein J.L., Hassan M.Q., Gaur T., Zhang Y., MicroRNA control of bone formation and homeostasis, Nat. Rev. Endocrinol., 2012, 8, 212-227

[78] Taipaleenmäki H., Bjerre Hokland L., Chen L., Kauppinen S., Kassem M., Mechanisms in endocrinology: micro-RNAs: targets for enhancing osteoblast differentiation and bone formation, Eur. J. Endocrinol., 2012, 166, 359-371

[79] Nieminen P., Genetic basis of tooth agenesis, J. Exp. Zool. Part B, 2009, 312B, 320-342

[80] Cao H., Wang J., Li X., Florez S., Huang Z., Venugopalan S.R., et al., MicroRNAs play a critical role in tooth development, J. Dent. Res., 2010, 89, 779-784

[81] Chou R.H., Yu Y.L., Hung M.C., The roles of EZH2 in cell lineage commitment, Am. J. Transl. Res., 2011, 3, 243-250

[82] Vire E., Brenner C., Deplus R., Blanchon L., Fraga M., Didelot C., The Polycomb group protein EZH2 directly controls DNA methylation, Nature, 2006, 439, 871-874

[83] Ronan J.L., Wu W., Crabtree G.R., From neural development to cognition: unexpected roles for chromatin, Nat. Rev. Genet., 2013, 14, 347-359

[84] Hirabayashi Y., Suzki N., Tsuboi M., Endo T.A., Toyoda T., Shinga J., et al., Polycomb limits the neurogenic competence of neural precursor cells to promote astrogenic fate transition, Neuron, 2009, 63, 600-613

[85] Schwarz D., Varum S., Zemke M., Schöler A., Baggiolini A., Draganova K., et al., Ezh2 is required for neural crest-derived cartilage and bone formation, Development, 2014, 141, 867-877

[86] Wei Y., Chen Y.H., Li L.Y., Lang J., Yeh S.P., Shi B., et al., CDK1-dependent phosphorylation of EZH2 suppresses methylation of H3K27 and promotes osteogenic differentiation of human mesenchymal stem cells, Nat. Cell Biol., 2011, 13, 87-94

[87] Yang D.C., Tsay H.J., Lin S.Y., Chiou S.H., Li M., Chang T.J., Hung S.C., cAMP/PKA regulates osteogenesis, adipogenesis and ratio of RANKL/OPG mRNA expression in mesenchymal stem cells by suppressing leptin, PLoS One, 2008, 3, e1540

[88] Tatton-Brown K., Hanks S., Ruark E., Zachariou A., Duarte Sdel V., Ramsay E., et al., Germline mutations in the oncogene EZH2 cause Weaver syndrome and increased human height, Oncotarget, 2011, 2, 1127-1133

[89] Gibson W.T., Hood R.L., Zhan S.H., Bulman D.E., Fejes A.P., Moore R., et al., Mutations in EZH2 cause Weaver syndrome, Am. J. Hum. Genet., 2012, 90, 110-118

[90] Bialek P., Kern B., Yang X., Schrock M., Sosic D., Hong N., et al., A twist code determines the onset of osteoblast differentiation, Dev. Cell, 2004, 6, 423-435

[91] Lian J.B., Stein G.S., van Wijnen A.J., Stein J.L., Hassan M.Q., Gaur T., et al., MicroRNA control of bone formation and homeostasis, Nat. Rev. Endocrinol., 2012, 8, 212-227

[92] Song X., Wang S., Li L., New insights into the regulation of Axin function in canonical Wnt signaling pathway, Protein Cell, 2014, 5, 186-193

[93] Jho E.H., Zhang T., Domon C., Joo C.K., Freund J.N., Costantini F., Wnt/beta-catenin/Tcf signaling induces the transcription of Axin2, a negative regulator of the signaling pathway, Mol. Cell Biol., 2002, 22, 1172-1183

[94] Chia I.V., Costantini F., Mouse axin and axin2/conductin proteins are functionally equivalent in vivo, Mol. Cell Biol., 2005, 25, 4371-4376

[95] Lammi L., Arte S., Somer M., Jarvinen H., Lahermo P., Thesleff I., et al., Mutations in AXIN2 cause familial tooth agenesis and predispose to colorectal cancer, Am. J. Hum. Genet., 2004, 74, 1043-1050

[96] Ng D., Thakker N., Corcoran C.M., Donnai D., Perveen R., Schneider A., et al., Oculofaciocardiodental and Lenz microphthalmia syndromes result from distinct classes of mutations in BCOR, Nat. Genet., 2004, 36, 411-416

[97] Cai J., Kwak S., Lee J.M., Kim E.J., Lee M.J., Park G.H., Cho S.W., Jung H.S., Function analysis of mesenchymal Bcor in tooth development by using RNA interference, Cell Tissue Res., 2010, 341, 251-258

[98] Yoshida T., Miyoshi J., Takai Y., Thesleff I., Cooperation of nectin-1 and nectin-3 is required for normal ameloblast function and crown shape development in mouse teeth, Dev. Dynam., 2010, 239, 2558-2569

[99] Brauer A.U., Savaskan N.E., Kuhn H., Prehn S., Ninnemann O., Nitsch R., A new phospholipid phosphatase, PRG-1, is involved in axon growth and regenerative sprouting, Nat. Neurosci., 2003, 6, 572-578

[100] Savaskan N.E., Brauer A.U., Nitsch R., Molecular cloning and expression regulation of PRG-3, a new member of the plasticity-related gene family, Eur. J. Neurosci., 2004, 19, 212-220

[101] Brauer A.U., Nitsch R., Plasticity-related genes (PRGs/LRPs): a brain-specific class of lysophospholipid-modifying proteins, Biochim. Biophys. Acta, 2008, 1781, 595-600

[102] Tokumitsu H., Hatano N., Tsuchiya M., Yurimoto S., Fujimoto T., Ohara N., et al., Identification and characterization of PRG-1 as a neuronal calmodulin-binding protein, Biochem. J., 2010, 431, 81-91

[103] Trimbuch T., Beed P., Vogt J., Schuchmann S., Synaptic PRG-1 modulates excitatory transmission via lipid phosphate-mediated signaling, Cell, 2009, 138, 1222-1235

[104] Strauss U., Bräuer A.U., Current views on regulation and function of plasticity-related genes (PRGs/LPPRs) in the brain, Biochim. Biophys. Acta, 2013, 1831, 133-138

[105] Ni H., Jiang Y.W., Tao L.Y., Jin M.F., Wu X.R., ZnT-1, ZnT-3, CaMK II, PRG-1 expressions in hippocampus following neonatal seizure-induced cognitive deficit in rats, Toxicol. Lett., 2009, 184, 145-150

[106] Ni H., Jiang Y.W., Xiao Z.J., Tao L.Y., Jin M.F., Wu X.R., Dynamic pattern of gene expression of ZnT-1, ZnT-3 and PRG-1 in rat brain following flurothyl-induced recurrent neonatal seizures, Toxicol. Lett., 2010, 194, 86-93

[107] Ni H., Feng X., Xiao Z.J., Tao L.Y., Jin M.F., Dynamic pattern of gene expression of ZnT-4, caspase-3, LC3, and PRG-3 in rat cerebral cortex following flurothyl-induced recurrent neonatal seizures, Biol. Trace Elem. Res., 2011, 143, 1607-1615

[108] Bando S.Y., Alegro M.C., Amaro E. Jr., Silva A.V., Castro L.H., Wen H.T., et al., Hippocampal CA3 transcriptome signature correlates with initial precipitating injury in refractory mesial temporal lobe epilepsy, PLoS One, 2011, 6, 10, e26268

[109] Broggini T., Nitsch R., Savaskan N.E., Plasticity-related gene 5 (PRG5) induces filopodia and neurite growth and impedes lysophosphatidic acid- and nogo-A-mediated axonal retraction, Mol. Biol. Cell, 2010, 21, 521-537

[110] Xu L., Yin W., Xia J., Peng M., Li S., Lin S., Pei D., Shu X., An antiapoptotic role of sorting nexin 7 is required for liver development in zebrafish, Hepatology, 2012, 55, 1985-1993

[111] van Weering J.R., Verkade P., Cullen P.J., SNX-BAR proteins in phosphoinositide-mediated, tubular-based endosomal sorting, Semin. Cell Dev. Biol., 2010, 21, 371-380

[112] Teasdale R.D., Collins B.M., Insights into the PX (phox-homology) domain and SNX (sorting nexin) protein families: structures, functions and roles in disease, Biochem. J., 2012, 441, 39-59

[113] Bravo J., Karathanassis D., Pacold C.M., Pacold M.E., Ellson C.D., Anderson K.E., The crystal structure of the PX domain from p40(phox) bound to phosphatidylinositol 3-phosphate, Mol. Cell., 2001, 8, 829-839

[114] Haberg K., Lundmark R., Carlsson S., SNX18 is an SNX9 paralog that acts as a membrane tubulator in AP-1-positive endosomal trafficking, J. Cell Sci., 2008, 121, 1495-1505

[115] Drin G., Casella J.F., Gautier R., Boehmer T., Schwartz T.U., Antonny B., A general amphipathic alpha-helical motif for sensing membrane curvature, Nat. Struct. Mol. Biol., 2007, 14, 138-146

[116] Huotari J., Helenius A., Endosome maturation, EMBO J., 2011, 30, 3481-500

[117] Van Weering J.R.T., Verkade P., Cullen P.J., SNX–BAR-mediated endosome tubulation is co-ordinated with endosome maturation, Traffic, 2012, 13, 94-107

[118] Vervoort V.S., Viljoen D., Smart R., Suthers G., DuPont B.R., Abbott A., Schwartz C.E., Sorting nexin 3 (SNX3) is disrupted in a patient with a translocation t(6;13)(q21;q12) and microcephaly, microphthalmia, ectrodactyly, prognathism (MMEP) phenotype, J. Med. Genet., 2002, 39, 893-899

[119] Wang X., Zhao Y., Zhang X., Badie H., Zhou Y., Mu Y., et al., Loss of sorting nexin 27 contributes to excitatory synaptic dysfunction by modulating glutamate receptor recycling in Down's syndrome, Nat. Med., 2013, 19, 473-480

[120] Seelan R.S., Khalyfa A., Lakshmanan J., Casanova M.F., Parthasarathy R.N., Deciphering the lithium transcriptome: microarray profiling of lithium-modulated gene expression in human neuronal cells, Neuroscience, 2008, 151, 1184-1197

[121] Becker K., Di Donato N., Holder-Espinasse M., Andrieux J., Cuisset J.M., Vallée L., Plessis G., et al., De novo microdeletions of chromosome 6q14.1-q14.3 and 6q12.1-q14.1 in two patients with intellectual disability - further delineation of the 6q14 microdeletion syndrome and review of the literature, Eur. J. Med. Genet., 2012, 55, 490-497

[122] Huang H.S., YoonBecker K., Di Donato N., Holder-Espinasse M., Andrieux J., Cuisset J.M., Vallée L., Plessis G., et al., Snx14 regulates neuronal excitability, promotes synaptic transmission, and is imprinted in the brain of mice, PLoS One, 2014, 9, e98383

[123] Wilkinson L.S., Davies W., Isles A.R., Genomic imprinting effects on brain development and function, Nat. Rev. Neurosci., 2007, 8, 832-843

[124] Maenaka S., Hikichi T., Imai M.A., Minamoto T., Kawahara E., Loss of imprinting in IGF2 in colorectal carcinoma assessed by microdissection, Oncol. Rep., 2006, 15, 791-795

[125] Kamikihara T., Arima T., Kato K., Matsuda T., Kato H., Douchi T., et al., Epigenetic silencing of the imprinted gene ZAC by DNA methylation is an early event in the progression of human ovarian cancer, Int. J. Cancer, 2005, 115, 690-700

[126] Coolen M., Katz S., Bally-Cuif L., miR-9: a versatile regulator of neurogenesis, Front. Cell Neurosci., 2013, 20, 7, 220

[127] Akerblom M., Sachdeva R., Jakobsson J., Functional studies of microRNAs in neural stem cells: problems and perspectives, Front. Neurosci., 2012, 7, 6-14

[128] Lau P., de Strooper B., Dysregulated microRNAs in neurodegenerative disorders, Semin. Cell Dev. Biol., 2010, 21, 768-773

[129] van Kuilenburg A.B., Vreken P., Abeling N.G., Bakker H.D., Meinsma R., Van Lenthe H., et al., Genotype and phenotype in patients with dihydropyrimidine dehydrogenase deficiency, Hum. Genet., 1999,104, 1-9

[130] Amstutz U., Froehlich T.K., Largiadèr C.R., Dihydropyrimidine dehydrogenase gene as a major predictor of severe 5-fluorouracil toxicity, Pharmacogenomics, 2011, 12, 1321-1336

[131] Collie-Duguid E.S., Etienne M.C., Milano G., McLeod H.L., Known variant DPYD alleles do not explain DPD deficiency in cancer patients, Pharmacogenetics, 2000, 10, 217-223

[132] Maekawa K., Saeki M., Saito Y., Ozawa S., Kurose K., Kaniwa N., et al., Genetic variations and haplotype structures of the DPYD gene encoding dihydropyrimidine dehydrogenase in Japanese and their ethnic differences, J. Hum. Genet., 2007, 52, 804-819

[133] Froehlich T.K., Amstutz U., Aebi S., Joerger M., Largiadèr C.R., Clinical importance of risk variants in the dihydropyrimidine dehydrogenase gene for the prediction of early-onset fluoropyrimidine toxicity, Int. J. Cancer, 2015, 136, 730-739

[134] van Kuilenburg A.B.P., Dobritzsch D., Meinsma R., Haasjes J., Waterham H.R., et al., Novel disease-causing mutations in the dihydropyrimidine dehydrogenase gene interpreted by analysis of the three-dimensional protein structure, Biochemical J., 2002, 364, 157-163

[135] Loganayagam A., Arenas Hernandez M., Corrigan A., Fairbanks L., Lewis C.M., Harper P., et al., Pharmacogenetic variants in the DPYD, TYMS, CDA and MTHFR genes are clinically significant predictors of fluoropyrimidine toxicity, Brit. J. Cancer, 2013, 108, 2505-2515

[136] Terrazzino S., Cargnin S., Del Re M., Danesi R., Canonico P.L., Genazzani A.A., DPYD IVS14+1G>A and 2846A>T genotyping for the prediction of severe fluoropyrimidine-related toxicity: a meta-analysis, Pharmacogenomics, 2013, 14, 1255-1272

[137] Morel A., Boisdron-Celle M., Fey L., Soulie P., Craipeau M.C., Traore S., Gamelin E., Clinical relevance of different dihydropyrimidine dehydrogenase gene single nucleotide polymorphisms on 5-fluorouracil tolerance, Mol. Cancer Ther., 2006, 11, 2895-2904

[138] Schwab M., Zanger U.M., Marx C., Schaeffeler E., Klein K., Dippon J., et al., German 5-FU Toxicity Study Group, Role of genetic and nongenetic factors for fluorouracil treatment-related severe toxicity: a prospective clinical trial by the German 5-FU Toxicity Study Group, J. Clin. Oncol., 2008, 26, 2131-2138

[139] Amstutz U., Farese S., Aebi S., Largiadèr CR., Dihydropyrimidine dehydrogenase gene variation and severe 5-fluorouracil toxicity: a haplotype assessment, Pharmacogenomics, 2009, 10, 931-944

[140] van Kuilenburg A.B., Meijer J., Mul A.N., Meinsma R., Schmid V., Dobritzsch D., et al., Intragenic deletions and a deep intronic mutation affecting pre-mRNA splicing in the dihydropyrimidine dehydrogenase gene as novel mechanisms causing 5-fluorouracil toxicity, Hum. Genet., 2010, 128, 529-538

[141] Collie-Duguid E.S., Etienne M.C., Milano G., McLeod H.L., Known variant DPYD alleles do not explain DPD deficiency in cancer patients, Pharmacogenetics, 2000, 10, 217-223

[142] van Kuilenburg A.B., van Lenthe H., Tromp A., Veltman P.C., van Gennip A.H., Pitfalls in the diagnosis of patients with a partial dihydropyrimidine dehydrogenase deficiency, Clin. Chem., 2000, 46, 9-17

[143] Noguchi T., Tanimoto K., Shimokuni T., Ukon K., Tsujimoto H., Fukushima M., et al., Aberrant methylation of DPYD promoter, DPYD expression, and cellular sensitivity to 5-fluorouracil in cancer cells, Clin. Cancer Res., 2004, 10, 7100-7107

[144] Amstutz U., Farese S., Aebi S., Largiadèr C., Hypermethylation of the DPYD promoter region is not a major predictor of severe toxicity in 5-fluorouracil based chemotherapy, J. Exp. Clin. Canc. Res, 2008, 27, 54

[145] Savva-Bordalo J., Ramalho-Carvalho J., Pinheiro M., Costa V.L., Rodrigues A., Dias P.C., et al., Promoter methylation and large intragenic rearrangements of DPYD are not implicated in severe toxicity to 5-fluorouracil-based chemotherapy in gastrointestinal cancer patients, BMC Cancer, 2010, 10, 470

[146] van Kuilenburg A.B., Meijer J., Mul A.N., Hennekam R.C., Hoovers J.M., de Die-Smulders C.E., et al., Analysis of severely affected patients with dihydropyrimidine dehydrogenase deficiency reveals large intragenic rearrangements of DPYD and a de novo interstitial deletion del(1)(p13.3p21.3), Hum. Genet., 2009, 125, 581-590

[147] Meng L., Person R.E., Beaudet A.L., Ube3a-ATS is an atypical RNA polymerase II transcript that represses the paternal expression of Ube3a, Hum. Mol. Genet., 2012, 21, 3001-3012

[148] Meng L., Person R.E., Huang W., Zhu P.J., Costa-Mattioli M., Beaudet A.L., Truncation of Ube3a-ATS unsilences paternal Ube3a and ameliorates behavioral defects in the Angelman syndrome mouse model, PLoS Genet., 2013, 9, e1004039

[149] Powell W.T., Coulsona R.L., Gonzalesa M.L., Crarya F.K., Wonga S.S., Adamsa S., et al., R-loop formation at Snord116 mediates topotecan inhibition of Ube3a-antisense and allele-specific chromatin decondensation, Proc. Nat. Acad. Sci. USA, 110, 13938-13943

[150] Faghihi M.A., Wahlestedt C., Regulatory roles of natural antisense transcripts, Nat. Rev. Mol. Cell Biol., 2009, 10, 637-643

[151] Magistri M., Faghihi M.A., St Laurent G., Wahlestedt C., Regulation of chromatin structure by long noncoding RNAs: focus on natural antisense transcripts, Trends Genet., 2012 28, 389-396

[152] Clark B., Blackshaw S., Long noncoding RNA-dependent transcriptional regulation in neuronal development and disease, Front. Genet., 2014, 5, 164

[153] Werner A., Cockell S., Falconer J., Carlile M., Alnumeir S., Robinson J., Contribution of natural antisense transcription to an endogenous siRNA signature in human cells, BMC Genomics 2014, 15, 1-12

[154] Chen J., Sun M., Kent W.J., Huang X., Xie H., Wang W., Zhou G., Shi R.Z., Rowley J.D., Over 20% of human transcripts might form sense-antisense pairs, Nucleic Acids Res., 2004, 32, 4812-4820

[155] Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N., et al., The transcriptional landscape of the mammalian genome, Science, 2005, 309, 1559-1563

[156] Katayama S., Tomaru Y., Kasukawa T., Waki K., Nakanishi M., Nakamura M., et al., RIKEN Genome Exploration Research Group; Genome Science Group (Genome Network Project Core Group); FANTOM Consortium, Antisense transcription in the mammalian transcriptome, Science, 2005, 309, 1564-1566

[157] Galante P.A., Vidal D.O., De Souza J.E., Camargo A.A., De Souza, S.J., Sense antisense pairs in mammals: functional and evolutionary considerations, Genome Biol., 2007, 8, R40

[158] Sun M., Hurst L.D., Carmichael G.G., and Chen J., Evidence for a preferential 1580 targeting of 3′-UTRs by cis-encoded natural antisense transcripts, Nucleic Acids Res., 2005, 33, 5533-5543

[159] Core L.J., Waterfall J.J., Lis J.T., Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters, Science, 2008, 322, 1845-1848

[160] Seila A.C., Calabrese J.M., Levine S., Yeo G.W., Rahl P.B., Flynn R.A., et al., Divergent transcription from active promoters, Science, 2008, 322, 1849-1851

[161] Vanhee-Brossollet C., Vaquero C., Do natural antisense transcripts make sense in eukaryotes?, Gene, 1998, 211, 1-9

[162] Alfano G., Vitiello C., Caccioppoli C., Caramico T., Carola A., Szego M.J., et al., Natural antisense transcripts associated with genes involved in eye development, Hum. Mol. Genet., 2005, 14, 913-923

[163] Carlile M., Swan D., Jackson K., Preston-Fayers K., Ballester B., Flicek P., et al., Strand selective generation of endo-siRNAs from the Na/phosphate transporter gene Slc34a1 in murine tissues, Nucleic Acids Res., 2009, 37, 2274-2282

[164] Kim T.H., Abdullaev Z.K., Smith A.D., Ching K.A., Loukinov D.I., Green R.D., et al., Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome, Cell, 2007, 128, 1231-1245

[165] Gaszner M., Felsenfeld G., Insulators: exploiting transcriptional and epigenetic mechanisms, Nat. Rev. Genet., 2006, 7, 703-713

[166] Pasmant E., Laurendeau I., Héron D., Vidaud M., Vidaud D., Bièche I., Characterization of a germ-line deletion, including the entire INK4/ARF locus, in a melanoma-neural system tumor family: identification of ANRIL, an antisense noncoding RNA whose expression coclusters with ARF, Cancer Res., 2007, 67, 3963-3969

[167] Yu W., Gius D., Onyango P., Muldoon-Jacobs K., Karp J., Feinberg A.P., Cui H., Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA,. Nature, 2008, 451, 202-206

[168] Popov N., Gil J., Epigenetic regulation of the INK4b-ARF-INK4a locus: in sickness and in health, Epigenetics, 2010, 5, 685-690

[169] Verona R.I., Mann M.R., Bartolomei M.S., Genomic imprinting: intricacies of epigenetic regulation in clusters, Ann. Rev. Cell Dev. Bi., 2003, 19, 237-259

[170] Nagano T., Mitchell J.A., Sanz L.A., Pauler F.M., Ferguson-Smith A.C., Feil R., et al., The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin, Science, 2008, 322, 1717-1720

[171] Pandey R.R., Mondal T., Mohammad F., Enroth S., Redrup L., Komorowski J., et al., Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation, Mol. Cell, 2008, 32, 232-246

[172] Terranova R., Yokobayashi S., Stadler M.B., Otte A.P., Van Lohuizen M., Orkin S.H., et al., Polycomb group proteins Ezh2 and Rnf2 direct genomic contraction and imprinted repression in early mouse embryos, Dev. Cell, 2008, 15, 668-679

[173] Wan L.B., and Bartolomei M.S., Regulation of imprinting in clusters: noncoding RNAs versus insulators, Adv. Genet., 2008, 61, 207-223

[174] Mohammad F., Mondal T., Kanduri, C., Epigenetics of imprinted long noncoding RNAs, Epigenetics, 2009, 4, 277-286

[175] Chamberlain S.J., Brannan C.I., The Prader-Willi syndrome imprinting center activates the paternally expressed murine Ube3a antisense transcript but represses paternal Ube3a, Genomics, 2001, 73, 316-322

[176] Rougeulle C., Cardoso C., Fontes M., Colleaux L., Lalande M., An imprinted antisense RNA overlaps UBE3A and a second maternally expressed transcript, Nat. Genet., 1998, 19, 15-16

[177] Dermitzakis E.T., Reymond A., Antonarakis S.E., Conserved non-genic sequences - an unexpected feature of mammalian genomes, Nat. Rev. Genet., 2005, 6, 151-157

[178] Louro R., El-Jundi T., Nakaya H.I., Reis E.M., Verjovski-Almeida, S., Conserved tissue expression signatures of intronic noncoding RNAs transcribed from human and mouse loci, Genomics, 2008, 92, 18-25

[179] St Laurent G., Shtokalo D., Tackett M.R., Yang Z., Eremina T., Wahlestedt C., et al., Intronic RNAs constitute the major fraction of the non-coding RNA in mammalian cells, BMC Genomics, 2012, 13, 504

[180] Nakaya H.I., Amaral P.P., Louro R., Lopes A., Fachel A.A., Moreira Y.B., et al., Genome mapping and expression analyses of human intronic noncoding RNAs reveal tissue-specific patterns and enrichment in genes related to regulation of transcription, Genome Biol., 2007, 8, R43

[181] Dinger M.E., Amaral P.P., Mercer T.R., Pang K.C., Bruce S.J., Gardiner B.B., et al., Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation, Genome Res., 2008, 18, 1433-1445

[182] Mercer T.R., Dinger, M.E., Sunkin, S.M., Mehler, M.F., and Mattick, J.S., Specific expression of long noncoding RNAs in the mouse brain, Proc. Nat. Acad. Sci. USA, 2008, 105, 716-721

[183] Rinn J.L., Euskirchen G., Bertone P., Martone R., Luscombe N.M., Hartman S., et al., The transcriptional activity of human Chromosome 22, Gene. Dev., 2003, 17, 529-540

[184] Bertone P., Stolc V., Royce T.E., Rozowsky J.S., Urban A.E., Zhu X., et al., Global identification of human transcribed sequences with genome tiling arrays, Science, 2004, 306, 2242-2246

[185] Kampa D., Cheng J., Kapranov P., Yamanaka M., Brubaker S., Cawley S., et al., Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22, Genome Res., 2004, 14, 331-342

[186] Guil S., Soler M., Portela A., Carrere J., Fonalleras E., Gomez A., et al., Intronic RNAs mediate EZH2 regulation of epigenetic targets, Nat. Struct. Mol. Biol., 2012, 19, 664-670

[187] Guil S., Esteller, M., Cis-acting noncoding RNAs: friends and foes, Nat. Struct. Mol. Biol., 2012, 19, 1068-1075

[188] Saif M.W., Ezzeldin H., Vance K., Sellers S., Diasio R.B., DPYD*2A mutation: the most common mutation associated with DPD deficiency, Cancer Chemoth. Pharm., 2007, 60, 503-507

[189] Rinn J.L., Kertesz M., Wang J.K., Squazzo S.L., Xu X., Brugmann S.A., et al., Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs, Cell, 2007, 129,1311-1323

[190] Ng K., Pullirsch D., Leeb M., Wutz A., Xist and the order of silencing, EMBO Rep., 2007, 8, 34-39

[191] Dinger M.E., Pang K.C., Mercer T.R., Mattick J.S., Differentiating protein-coding and noncoding RNA: challenges and ambiguities, PLoS Comput. Biol., 2008, 4, e1000176

[192] Khalil A.M., Guttman M., Huarte M., Garber M., Raj A., Rivea Morales D., et al., Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression, Proc. Natl. Acad. Sci. USA, 2009, 106, 11667-11672

[193] Mattick J.S., Amaral P.P., Dinger M.E., Mercer T.R., Mehler M.F., RNA regulation of epigenetic processes, Bioessays, 2009, 31, 51-59

[194] Redrup L., Branco M.R., Perdeaux E.R., Krueger C., Lewis A., Santos F., et al., The long noncoding RNA Kcnq1ot1 organises a lineage-specific nuclear domain for epigenetic gene silencing, Development, 2009, 136, 525-530

[195] Hormozian F., Schmitt J.G., Sagulenko E., Schwab M., Savelyeva L., FRA1E common fragile site breaks map within a 370 kilobase pair region and disrupt the dihydropyrimidine dehydrogenase gene (DPYD), Cancer Lett., 2007, 246, 82-91

[196] Fungtammasan A., Walsh E., Chiaromonte F., Eckert K.A., Makova K.D., A genome-wide analysis of common fragile sites: what features determine chromosomal instability in the human genome?, Genome Res., 2012, 22, 993-1005

[197] Lukusa T., Fryns J.P., Human chromosome fragility, Biochim. Biophys. Acta, 2008, 1779, 3-16

[198] Mrasek K., Schoder C., Teichmann A-C., Behr K., Franze B., Wilhelm K., et al., Global screening and extended nomenclature for 230 aphidicolin-inducible fragile sites, including 61 yet unreported ones, Int. J. Oncol., 2010, 36, 929-940

[199] Zlotorynski E., Rahat A., Skaug J., Ben-Porat N., Ozeri E., Hershberg R., et al., Molecular basis for expression of common and rare fragile sites, Mol. Cell. Biol., 2003, 23, 7143-7151

[200] Travers A.A., The structural basis of DNA flexibility, Philos. Trans. A. Math. Phys. Eng. Sci., 2004, 362, 1423-1438

[201] Zhang H., Freudenreich C.H., An AT-rich sequence in human common fragile site FRA16D causes fork stalling and chromosome breakage in S. cerevisiae, Mol. Cell, 2007, 27, 367-379

[202] Shah S.N., Opresko P.L., Meng X., Lee M.Y.W.T., Eckert K.A., DNA structure and the Werner protein modulate human DNA polymerase d-dependent replication dynamics within the common fragile site FRA16D, Nucleic Acids Res., 2010, 38, 1149-1162

[203] Dillon L.W., Burrow A.A., Wang Y-H., DNA instability at chromosomal fragile sites in cancer, Curr. Genomics, 2010, 11, 326-337

[204] Durkin S.G., Glover T.W., Chromosome fragile sites, Annu. Rev. Genet., 2007, 41, 169-192

[205] Debatisse M., Le Tallec B., Letessier A., Dutrillaux B., Brison O., Common fragile sites: mechanisms of instability revisited, Trends Genet., 2012, 28, 22-32

[206] Gross E., Busse B., Riemenschneider M., Neubauer N., Seck K., Klein HG., et al., Strong association of a common dihydropyrimidine

dehydrogenase gene polymorphism with fluoropyrimidine-related toxicity in cancer patients, PLoS One, 2008, 3, e4003

[207] Ried K., Finnis M., Hobson L., Mangelsdorf M., Dayan S., Nancarrow J.K., et al., Common chromosomal fragile site FRA16D sequence: Identification of the FOR gene spanning FRA16D and homozygous deletions and translocation breakpoints in cancer cells, Hum. Mol. Genet., 2000, 9, 1651-1663

[208] Ragland R.L., Glynn M.W., Arlt M.F., Glover T.W., Stably transfected common fragile site sequences exhibit instability at ectopic sites, Gene Chromosome Canc., 2008, 47, 860-872

[209] Lander E.S., Linton L.M., Birren B., Nusbaum C., Zody M.C., Baldwin J., et al., Initial sequencing and analysis of the human genome, Nature, 2001, 409, 860-921

[210] Craig N.L., Eickbush T.H., Voytas D.F., Welcome to mobile DNA, Mob. DNA, 2010, 1, 1

[211] Pace J.K., Feschotte C., The evolutionary history of human DNA transposons: Evidence for intense activity in the primate lineage, Genome Res., 2007, 17, 422-432

[212] Belancio V.P., Hedges D.J., Deininger P., Mammalian non-LTR retrotransposons: for better or worse, in sickness and in health, Genome Res., 2008, 18, 343-358

[213] Belancio V.P., Deininger P.L., Roy-Engel A.M., LINE dancing in the human genome: transposable elements and disease, Genome Med., 2009, 1, 97

[214] Dombroski B.A., Mathias S.L., Nanthakumar E., Scott A.F., Kazazian H.H., Isolation of an active human transposable element, Science, 1991, 254, 1805-1808

[215] Miki Y., Nishisho I., Horii A., Miyoshi Y., Utsunomiya J., Kinzler K.W., et al., Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer, Cancer Res., 1992, 52, 643-645

[216] Pickeral O.K., Makalowski W., Boguski M.S., Boeke J.D., Frequent human genomic DNA transduction driven by LINE-1 retrotransposition, Genome Res., 2000, 10, 411-415

[217] Han K., Lee J., Meyer T.J., Remedios P., Goodwin L., Batzer,M.A., L1 recombination-associated deletions generate human genomic variation, Proc. Natl. Acad. Sci. USA, 2008, 105, 19366-19371

[218] Robberecht C., Voet T., Zamani Esteki M., Nowakowska B.A., Vermeesch J.R., Nonallelic homologous recombination between retrotransposable elements is a driver of de novo unbalanced translocations, Genome Res., 2013, 23, 411-418

[219] Gasior S.L., Wakeman T.P., Xu B., Deininger P.L., The human LINE-1 retrotransposon creates DNA double-strand breaks, J. Mol. Biol., 2006, 357, 1383-1393

[220] Kines K.J., Sokolowski M., deHaro D.L., Christian C.M., Belancio V.P., Potential for genomic instability associated with retrotranspositionally-incompetent L1 loci, Nucleic Acids Res., 2014, 42, 10488-10502

[221] Kines K.J., Belancio V.P., Expressing genes do not forget their LINEs: transposable elements and gene expression, Front. Biosci. (Landmark Ed.), 2012, 17, 1329-1344

[222] Belancio V.P., Roy-Engel A.M., Pochampally R.R., Deininger P., Somatic expression of LINE-1 elements in human tissues, Nucleic Acids Res., 2010, 38, 3909-3922

[223] Belancio V.P., Roy-Engel A.M., Deininger P.L., All y'all need to know 'bout retroelements in cancer, Semin. Cancer Biol., 2010, 20, 200-210

[224] Wallace N.A., Belancio V.P., Faber Z., Deininger P., Feedback inhibition of L1 and alu retrotransposition through altered double strand break repair kinetics, Mob. DNA, 2010, 1, 22

[225] Kramerov D.A., Vassetzky NS., Short retroposons in eukaryotic genomes, Rev. Cytol., 2005, 247, 165-221

[226] Ostertag E.M., Goodier J.L., Zhang Y., Kazazian H.H., SVA elements are nonautonomous retrotransposons that cause disease in humans, Am. J. Hum. Genet., 2003, 73, 1444-1451

[227] Skowronski J., Singer M.F., Expression of a cytoplasmic LINE-1 transcript is regulated in a human teratocarcinoma cell line, Proc. Natl. Acad. Sci. USA, 1985, 82, 6050-6054

[228] Perepelitsa-Belancio V., Deininger P., RNA truncation by premature polyadenylation attenuates human mobile element activity, Nat. Genet., 2003, 35, 363-366

[229] Belancio V.P., Hedges D.J., Deininger P., LINE-1 RNA splicing and influences on mammalian gene expression, Nucleic Acids Res., 2006, 34, 51512-1521

[230] Lavie L., Maldener E., Brouha B., Meese E.U., Mayer J., The human L1 promoter: variable transcription initiation sites and a major impact of upstream flanking sequence on promoter activity, Genome Res., 2004, 14, 2253-2260

[231] Swergold G.D., Identification, characterization, and cell specificity of a human LINE- 1 promoter, Mol. Cell Biol., 1990, 10, 6718-6729

[232] Severynse D.M., Hutchison C.A., Edgell M.H., Identification of transcriptional regulatory activity within the 5' A-type monomer sequence of the mouse LINE-1 retroposon, Mamm. Genome, 1992, 2, 41-50

[233] Nigumann P., Redik K., Matlik K., Speek M., Many human genes are transcribed from the antisense promoter of L1 retrotransposon, Genomics, 2002, 79, 628-634

[234] Speek M., Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes, Mol. Cell Biol., 2001, 21, 1973-1985

[235] Kazazian H.H. Jr., Goodier J.L., LINE drive retrotransposition and genome instability, Cell, 2002, 110, 277-280

[236] Yang N., Kazazian H.H. Jr., L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells, Nat. Struct. Mol. Biol., 2006, 13, 763-771

[237] Wheelan S.J., Aizawa Y., Han J.S., Boeke J.D., Gene-breaking: a new paradigm for human retrotransposon-mediated gene evolution, Genome Res., 2005, 15, 1073-1078

[238] Hata K., Sakaki Y., Identification of critical CpG sites for repression of L1 transcription by DNA methylation, Gene, 1997, 189, 227-234

[239] Montoya-Durango D.E., Liu Y., Teneng I., Kalbfleisch T., Lacy M.E., Steffen M.C., Ramos K.S., Epigenetic control of mammalian LINE-

1 retrotransposon by retinoblastoma proteins, Mutat. Res., 2009, 665, 20-28

[240] Huda A., Jordan I.K., Epigenetic regulation of Mammalian genomes by transposable elements, Ann. NY Acad. Sci., 2009, 1178, 276-284

[241] Allen E., Horvath S., Tong F., Kraft P., Spiteri E., Riggs AD, Marahrens Y., High concentrations of long interspersed nuclear element sequence distinguish monoallelically expressed genes, Proc. Natl. Acad. Sci. USA, 2003, 100, 9940-9945

[242] Harris C.R., Dewan A., Zupnick A., Normart R., Gabriel A., Prives C., Levine A.J., Hoh J., p53 responsive elements in human retrotransposons, Oncogene, 2009, 28, 3857-3865

[243] Morales J.F., Snow E.T., Murnane J.P., Environmental factors affecting transcription of the human L1 retrotransposon, I. Steroid hormone-like agents, Mutagenesis, 2002, 17, 193-200

[244] Morales J.F., Snow E.T., Murnane J.P., Environmental factors affecting transcription of the human L1 retrotransposon. II. Stressors, Mutagenesis, 2003, 18, 151-158

[245] Yang N., Zhang L., Zhang Y., Kazazian H.H., An important role for RUNX3 in human L1 transcription and retrotransposition, Nucleic Acids Res., 2003, 31, 4929-4940

[246] Beier F., Lammi M.J., Bertling W., von der Mark K., Transcriptional regulation of the human type X collagen gene expression, Ann. NY Acad. Sci., 1996, 785, 209-211

[247] Landry J.R., Medstrand P., Mager D.L., Repetitive elements in the 5′ untranslated region of a human zinc-finger gene modulate transcription and translation efficiency, Genomics, 2001, 76, 110-116

[248] Muotri A.R., Chu V.T., Marchetto M.C., Deng W., Moran J.V., Gage F.H, Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition, Nature, 2005, 435, 903-910

[249] Shephard E.A., Chandan P., Stevanovic-Walker M., Edwards M., Phillips I.R., Alternative promoters and repetitive DNA elements define the species-dependent tissue-specific expression of the FMO1 genes of human and mouse, Biochem. J., 2007, 406, 491-499

[250] Steel G., Lutz E.M., Characterisation of the mouse vasoactive intestinal peptide receptor type 2 gene, Vipr2, and identification of a polymorphic LINE-1-like sequence that confers altered promoter activity, J. Neuroendocrinol., 2007, 19, 14-25

[251] Matlik K., Redik K., Speek M., L1 antisense promoter drives tissue-specific transcription of human genes, J. Biomed. Biotechnol., 2006, 2006, 1-16

[252] Cruickshanks H.A., Tufarelli C., Isolation of cancer-specific chimeric transcripts induced by hypomethylation of the LINE-1 antisense promoter, Genomics, 2009, 94, 397-406

[253] Dunn C.A., van de Lagemaat L.N., Baillie G.J., Mager D.L., Endogenous retrovirus long terminal repeats as ready-to-use mobile promoters: the case of primate beta3GAL-T5, Gene, 2005, 364, 2-12

[254] Dewannieux M., Esnault C., Heidmann T., LINE-mediated retrotransposition of marked Alu sequences, Nat. Genet., 2003, 35, 41-48

[255] Belgnaoui S.M., Gosden R.G., Semmes O.J., Haoudi A., Human LINE-1 retrotransposon induces DNA damage and apoptosis in cancer cells, Cancer Cell Int., 2006, 6, 13-25

[256] Pierce A.J., Stark J.M., Araujo F.D., Moynahan M.E., Berwick M., Jasin,M., Double-strand breaks and tumorigenesis, Trends Cell Biol., 2001, 11, S52-S59

[257] Mills K.D., Ferguson D.O., Alt F.W., The role of DNA breaks in genomic instability and tumorigenesis, Immunol. Rev., 2003, 194, 77-95

[258] Longhese M.P., Mantiero D., Clerici M., The cellular response to chromosome breakage, Mol. Microbiol., 2006, 60, 1099-1108

[259] Ergün S., Buschmann C., Heukeshoven J., Dammann K., Schnieders F., Lauke H., et al., Cell type-specific expression of LINE-1 open reading frames 1 and 2 in fetal and adult human tissues, J. Biol. Chem., 2004, 279, 27753-27763

[260] Garcia-Perez J.L., Marchetto M.C., Muotri A.R., Coufal N.G., Gage F.H., O'Shea K.S., Moran J.V., LINE-1 retrotransposition in human embryonic stem cells, Hum. Mol. Genet., 2007, 16, 1569-1577

[261] Asch H.L., Eliacin E., Fanning T.G., Connolly J.L., Bratthauer G., Asch,B.B., Comparative expression of the LINE-1 p40 protein in human breast carcinomas and normal breast tissues, Oncol. Res., 1996, 8, 239-247

[262] Nangia-Makker P., Sarvis R., Visscher D.W., Bailey-Penrod J., Raz A., Sarkar F.H., Galectin-3 and L1 retrotransposons in human breast carcinomas, Breast Cancer Res. Treat., 1998, 49, 171-183

[263] Iskow R.C., McCabe M.T., Mills R.E., Torene S., Pittard W.S., Neuwald A.F., et al., Natural mutagenesis of human genomes by endogenous retrotransposons, Cell, 2010, 141, 1253-1261

[264] Solyom S., Ewing A.D., Rahrmann E.P., Doucet T., Nelson H.H., Burns M.B., et al., Extensive somatic L1 retrotransposition in colorectal tumors, Genome Res., 2012, 22, 2328-2338

[265] Lee E., Iskow R., Yang L., Gokcumen O., Haseley P., Luquette L.J., et al., Landscape of somatic retrotransposition in human cancers, Science, 2012, 337, 967-971

[266] Sen S.K., Han K., Wang J., Lee J., Wang H., Callinan P.A., et al., Human genomic deletions mediated by recombination between Alu elements, Am. J. Hum. Genet., 2006, 79, 41-53

[267] Medstrand P., van de Lagemaat L.N., Mager D.L., Retroelement distributions in the human genome: variations associated with age and proximity to genes, Genome Res., 2002, 12, 1483-1495

[268] Smit A.F., Riggs A.D., Tiggers and DNA transposon fossils in the human genome, Proc. Natl. Acad. Sci. USA, 1996, 93, 1443-1448

[269] Smit A.F., Interspersed repeats and other mementos of transposable elements in mammalian genomes, Curr. Opin. Genet. Dev., 1999, 9, 657-663

[270] Kapitonov V., Jurka J., The age of Alu subfamilies, J. Mol. Evol., 1996, 42, 59-65

[271] Lim J.K., Simmons, M.J., Gross chromosome rearrangements mediated by transposable elements in Drosophila melanogaster, Bioessays, 1994, 16, 269-275

[272] Caceres M., Ranz J.M., Barbadilla A., Long M., Ruiz A., Generation of a widespread Drosophila inversion by a transposable element,

Science, 1999, 285, 415-418

[273] Gray Y.H., It takes two transposons to tango: Transposable-element-mediated chromosomal rearrangements, Trends Genet., 2000, 16, 461-468

[274] Zhang J., Peterson T., 2004, Transposition of reversed Ac element ends generates chromosome rearrangements in maize, Genetics, 167, 1929-1937

[275] Froyen G., Belet S., Martinez F., Santos-Rebouças C.B., Declercq M., Verbeeck J., et al., Copy-number gains of HUWE1 due to replication- and recombination-based rearrangements, Am. J. Hum. Genet., 2012, 91, 252-64

[276] Ivics Z., Hackett P.B., Plasterk R.H., Izsvak Z., Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells, Cell, 1997, 91, 501-510

[277] Luft F.C., Sleeping Beauty jumps to new heights, Mol. Med., 2010, 88, 641-643

[278] Ivics Z., Izsvak Z., Transposons for gene therapy!, Current Gene Ther., 2006, 6, 593-607

[279] Wilson M.H., Coates C.J., George A.L. Jr., PiggyBac transposon-mediated gene transfer in human cells, Mol. Ther., 2007, 15, 139-145

[280] Hackett P.B., Largaespada D.A., Cooper L.J., A transposon and transposase system for human application, Mol. Ther., 2010, 18, 674-683

[281] Singh H., Figliola M.J., Dawson M.J., Olivares S., Zhang L., Yang G., et al., Manufacture of clinical-grade CD19-specific T cells stably expressing chimeric antigen receptor using Sleeping Beauty system and artificial antigen presenting cells, PLoS One, 2013, 8, e64138

[282] Carlson C.M., Largaespada D.A., Insertional mutagenesis in mice: new perspectives and tools, Nat. Rev. Genet., 2006, 568-580

[283] Song G., Cui Z., Mob Genet Elements, Novel strategies for gene trapping and insertional mutagenesis mediated by Sleeping Beauty transposon, Mobile Genetic Elements, 2013, 3, e26499

[284] Ranzani M., Annunziato S., Adams D.J., Montini E., Cancer gene discovery: exploiting insertional mutagenesis, Mol. Cancer Res., 2013, 11, 1141-1158

[285] Strand D.J., McDonald J.F., Copia is transcriptionally responsive to environmental stress, Nucleic Acids Res., 1985, 13, 4401-4410

[286] Batzer M.A., Deininger P.L., Alu repeats and human genomic diversity, Nat. Rev. Genet., 2002, 3, 370-379

[287] Deininger P. L., Batzer M. A., Alu repeats and human disease, Mol. Genet. Metab., 1999, 67, 183-193

[288] Konkel M.K., Batzer M.A., A mobile threat to genome stability: The impact of non-LTR retrotransposons upon the human genome, Semin. Cancer. Biol., 2010, 20, 211-221

[289] Bose P., Hermetz K.E., Conneely K.N., Rudd M.K., Tandem repeats and G-rich sequences are enriched at human CNV breakpoints, PLoS One, 2014, 9, e101607

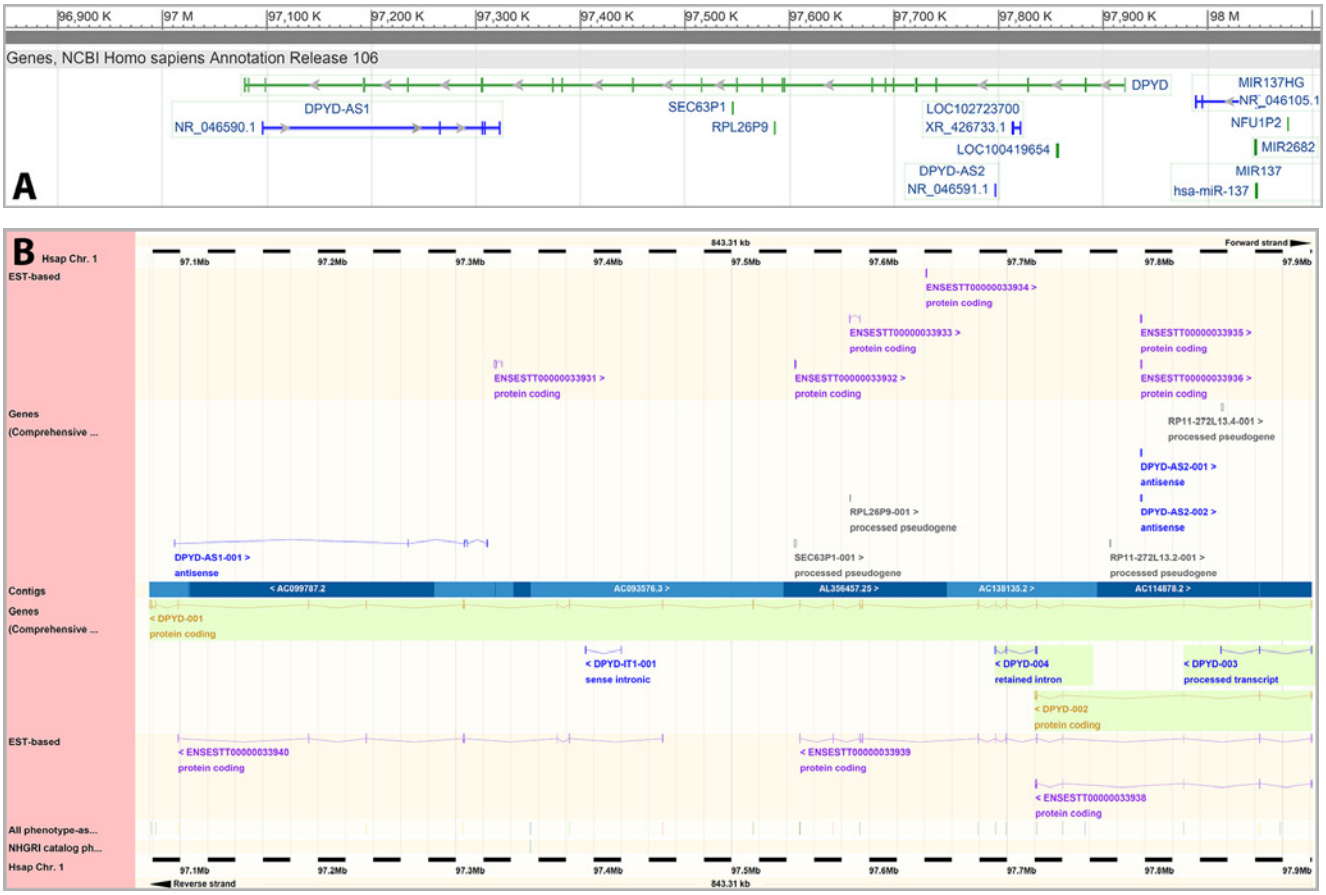
[290] Gebow D., Miselis N., Liber H. L., Homologous and nonhomologous recombination resulting in deletion: effects of p53 status, microhomology, and repetitive DNA length and orientation, Mol. Cell Biol., 2000, 20, 4028-4035

[291] Stenger J. E., Lobachev K.S., Gordenin D., Darden T.A., Jurka J., Resnick M.A., Biased distribution of inverted and direct Alus in the human genome: implications for insertion, exclusion, and genome stability. Genome Res., 2001, 11, 12-27

[292] Ade C., Roy-Engel A.M., Deininger P.L., Alu elements: an intrinsic source of human genome instability, Curr. Opin. Virol., 2013, 3, 639-45

[293] Boone P.M., Yuan B., Campbell I.M., Scull J.C., Withers M.A., Baggett B.C., et al., The Alu-rich genomic architecture of SPAST predisposes to diverse and functionally distinct disease-associated CNV alleles, Am. J. Hum. Genet., 2014, 95, 143-161

Supplementary material



Supplementary Figure 1 (A-B). Molecular organization of DPYD gene according to GRC38/hg38 Genome Annotation. A) NCBI Annotation Release 106 August 2014. Note the new lncRNA XR_426733.1/LOC102723700/ NC_018912.2; B) GRCh38 Ensembl Release August 2014. Note new 9 EST protein coding transcripts colored in purple. Fig. 1A and 2B are not drawn to scale.