

PROTEOME OF THE HUMAN CHROMOSOME 18: GENE-CENTRIC IDENTIFICATION OF TRANSCRIPTS, PROTEINS AND PEPTIDES

Addendum to the Roadmap: HEALTH ASPECTS

1. PROTEOMICS MEETS MEDICINE

At its very beginning, one of the goals of human proteomics became a disease biomarker discovery. Many works compared diseased and normal tissues and liquids to get diagnostic profiles by many proteomics methods. Of them, some cancer proteome profiling studies were considered too optimistic in terms of clinical applicability due to incorrect experimental design [Petricoin], thereby conferring the negative expectations from proteomics in translational medicine [Diamantidis, Nature]. The interlaboratory reproducibility of proteomics pipelines also was considered as a shortage in some papers, e.g. in the works of Bell et al [2009] who tested the proteome MS methods with 20-protein standard sample. These difficulties at the early stage of proteomics were partly caused by the fact that many attempts were mostly directed to the technique adjustment rather than to the clinically relevant result.

However, the recent advances in mass-spectrometry including the use of MRM to quantify peptides of proteome [Anderson Hunter 2006] made the community to have a view of cautious optimism on the problem of translation to medicine [Nilsson 2010]. A reproducibility problem stated in [Bell 2009] was shown to be mainly caused by the bioinformatics misinterpretation whereas the MS itself worked properly. The readiness of MRM-based platforms to the clinical use is illustrated by the attempt to pass FDA with the mock application which describes MS-based quantitation test for 10 proteins [Regnier FE 2010]. In its current state, the test has not got a clearance. However, some problems stated by

FDA representatives may be soon cleared as it is done with the mass-spectrometry for drug monitoring [Regnier FE 2010].

The gene-centric Human Proteome project was recently launched (9th World Congress of HUPO, Sydney, 2010; <http://www.hupo.org/research/hpp/soc/>). According to the technology roadmap of its Russian part (http://www.hupo.org/research/hpp/soc/RusRoadmap_Brief.Oct19_2010.pdf), the proteome expression from the human chromosome 18 should be studied. The project, at its pilot stage, is focused on MRM-based identification of proteins encoded by human chromosome 18 in model cells and selected tissues as well as on the development of technique to identify low-abundant proteins (on level less than 10^{-15} M). Following stages of the project contemplates the use of the proteomics of the chromosome 18 in clinical application. In this context, there is a need in the overview of medical background related to proteins encoded in the chromosome.

2. MONOGENIC HEREDITARY DISEASES LINKED WITH THE CHROMOSOME 18.

Some monogenic diseases are discovered that are associated with the deficiency of 18th chromosome-encoded proteins. Besides, there are known some chromosomal aberrations such as a trisomy, usually lethal Edwards syndrome which has 1:6000 frequency. A monosomy and a tetrasomy of the chromosome also are known and they occur very rarely. Effects of such disorders on a protein level are too complicated to be considered in scope of the gene-centric project. This section will be focused on the participation of several protein-coded genes in hereditary diseases.

Niemann-Pick disease, type C is a lysosomal storage disease with an autosomal dominant inheritance which occurs in approximately 1:120,000 people [Vanier, 2010]. 95% of the disease is caused by the mutations in *NPC1* gene (O15118¹). This gene encodes a large cell membrane protein which is involved in lipid transport. Diverse clinical manifestations of the disease with progressing neurological symptoms are responsible for the disability and caused by the lipid storage in tissues.

Erythropoietic protoporphyria (EPP) is a disease caused by the disordered heme synthesis with storage of protoporphyrins in erythrocytes, skin and liver [Lecha et al, 2009]. The disease strikes between 1:75,000 and 1:200,000 people and caused by mutations in the mitochondrial ferrochelatase *FECH* gene (P22830). It may have both autosomal dominant and recessive inheritance and, usually, is not life-threatening despite the painful skin symptoms of photosensitivity.

Transthyretin (also known as prealbumin) encoded by *TTR* gene (P02766) on 18q12.1 is a serum and cerebrospinal fluid carrier of the thyroid hormone thyroxine (T4) and vitamin A (retinol). Due to point mutations of *TTR* gene (V30M and V122I) TTR can abnormally accumulate as amyloid β -fibrils in tissues such as the heart, kidneys, nerves, and intestine causing neurodegeneration and organ failure [Saraiva, 2001]. Those autosomal dominant disorders are called **familial amyloid polyneuropathy (FAP)** and **familial amyloid cardiomyopathy (FAC)**. It should be mentioned that wild type protein also can deposit as amyloid causing senile systemic amyloidosis which affects over 25% of the population over age 80 [Tanskanen et al, 2008].

¹ The UniProt accession numbers of corresponding protein, as well as here and after (www.uniprot.org).

Other hereditary syndromes caused by mutations in chromosome 18 protein-coding genes are very rare and not more than 150 cases are described for each. Among them, a gene mutations of *TCF4* neuronal transcription factor (P15884) yields the **Pitt-Hopkins syndrome**. It is an autosomal dominant syndrome with mental retardation and serious respiratory failures [Amiel et al, 2007]. Mutations of *AFG32* gene encoding AFG3-like protein 2 (Q9Y4W6) are the cause of **spinocerebellar ataxia type 28** (SCA28) which is a very rare autosomal dominant ataxia with a slow progressive course [Maltecca et al, 2009]. AFG3-like protein 2 of itself is an ATP-dependent protease involved in axonal development. Besides, one of the types of **glucocorticoid deficiency, type 1** (GCCD1) is related to mutations in *MC2R* gene encoding the adrenocorticotrophic hormone receptor (Q01718). This autosomal recessive hormone deficiency is very rare and is described in about 150 subjects [Chung et al, 2008]. Further, one of many genes of holoprosencephaly, a disease with failure of brain hemisphere development, is *TGIF* encoding a homeobox protein TGIF1 transcription factor (Q15583). The TGIF-encoded disease is classified as **holoprosencephaly type 4** [El-Jaick et al, 2007]. Finally, the gene of transcriptional cofactor *SMAD4* (Q13485) is associated with autosomal dominant **juvenile polyposis** [Gallione et al, 2010]. This rare syndrome caused by *SMAD4* mutation has also enhanced risk of gastrointestinal tumors (for link of *SMAD4* to tumors see below a Cancer section). Defect in *LMAN1* gene (lectin, mannose-binding 1, also known as protein ERGIC-53, P49257) leads to **combined deficiency of factor V-factor VIII**, a rare, autosomal recessive disorder in which both coagulation factors V and VIII are diminished [Segal et al, 2004].

Selected hereditary syndromes were mapped on the chromosome 18 without exact information about change in protein-coding gene structure or they were mapped in non-coding regions of the chromosome. Among them, a rare **Seckel syndrome** [Borglum et al, 2001] and an autosomal dominant form of

obesity [Loos et al, 2008] should be mentioned. Linkage loci of 2 neurological autosomal dominant diseases, **torsion dystonia 7** [Leube et al, 1996] and **myoclonus dystonia** [Grimes et al, 2002], and of one autosomal recessive, **neurosensory deafness** [Mir et al, 2005], were also mapped on the chromosome without gene reference. The data on hereditary syndromes related to the protein-coding genes of the chromosome 18 is summarized in Table 1.

Table 1. Monogenic hereditary diseases linked with the chromosome 18

Location on the chromosome	Gene or region	Disease	Reference
18q11-q12	<i>NPC1</i>	Niemann-Pick disease, type C	Vanier, 2010
18q21.3	<i>FECH</i>	Erythropoietic protoporphyria	Lecha et al, 2009
18q21.1	<i>TCF4</i>	Pitt-Hopkins syndrome	Amiel et al, 2007
18q12.1	<i>TTR</i>	Familial amyloidosis	Saraiva, 2001
18p11	<i>AFG3L2</i>	Spinocerebellar ataxia type 28	Maltecca et al, 2009
18q11.1	<i>MC2R</i>	Glucocorticoid deficiency, type 1	Chung et al, 2008
18p11.3	<i>TGIF</i>	Holoprosencephaly, type 4	El-Jaick et al, 2007
18q21.1	<i>SMAD4</i>	Juvenile polyposis	Gallione et al, 2010
18p11.31-q11.2	<i>SCKL2</i> locus, trans-centromeric region	Seckel syndrome	Borglum et al, 2001

18q21.32	<i>rs 17782313</i> region (188 kb from <i>MC4R</i> gene)	Obesity, autosomal dominant	Loos et al, 2008
18p	<i>DYT7</i> region	Torsion dystonia 7	Leube et al, 1996
18p11.3	<i>DYT15</i> region	Myoclonus dystonia	Grimes et al, 2002
18q21.3-q22	<i>LMAN1</i>	Combined deficiency of factor V-factor VIII	Segal et al, 2004
18p11.32-11.31	<i>DFNB46</i> region	Neurosensory deafness	Mir et al, 2005

3. CANCER GENETICS OF THE CHROMOSOME 18.

Recent advances in cancer genetics are boosted by genome-wide studies of selected tumors and even cancer cell clones. Therefore mutations that cause or increase the risk of cancer and somatic mutation that occur during carcinogenesis may be considered separately. However, the presence of such germline and/or somatic mutations in one gene is usual in different types of cancer. Obviously, the knowledge about expression and level of cancer-related proteins obtainable in scope of Human Proteome Project may help to discover new biomarkers and drug target candidates.

SMAD proteins. SMADs (Similar to Mothers Against Dectapentaplegic, *Drosophila* gene) represent a group of proteins that form a complex carrying out a transcriptional regulation of TNF β pathway. Two receptor-dependent SMADs exist in the cell. One of them is SMAD2 (Q15796) which is encoded by the chromosome 18. *SMAD2* mutations occur in the colon and lung carcinoma tissues [Slattery et al, 2010a]. Gene of the cofactor SMAD4 (Q13485) is also located on the chromosome of interest. Somatic deletions of *SMAD4* gene are associated with

pancreatic cancer [Blackford et al, 2009]. In the contrary, SMAD7 (O15105) protein inhibits the complex of other SMADs and decreases the TNF β signaling. Due to proliferative trend of the latter, SMAD7 in its mutated state is an oncogene. Germline mutations in SMAD7 are described that elevate the risk of colorectal cancer [Slattery et al, 2010b]. In this cancer, SMAD7 somatic deletions also described [Boulay et al, 2003].

BCL-2 protein. B-cell lymphoma-2 (*BCL2*, P10415) is a well studied oncogene which is a title member for large family of pro-apoptotic and anti-apoptotic proteins. BCL-2 is a membrane molecule acting in the mitochondrial membrane. It prevents apoptosis when expressed. A classical chromosomal translocation (14;18) locates *BCL2* under a strong immunoglobulin promoter from chromosome 14 providing a B-cell lymphoma. This mutation may be both germline and somatic [Leich et al, 2009]. Moreover, the *BCL2* mutations are found in some solid cancers.

Translocations involving other chromosome 18 regions are also observed for **synovial sarcoma** and **T-cell acute lymphoblastic leukemia (T-ALL)**. Synovial sarcomas (SS) are infrequent and morphologically heterogeneous soft tissue sarcomas. In about 95% of synovial sarcomas, chromosomal translocation t(X;18)(p11;q11) was found which results in fusion of *SYT* gene (also known as SSXT; Synovial sarcoma, translocated to X chromosome, Q15532) with one of the members of gene family *SSX* (*SSX1*, *SSX2* or rarely *SSX4*) generating novel chimeric oncogene *SYT/SSX* which acts as transcription coactivator [dos Santos et al, 2001]. A *NUP98-SETBP1* gene fusion resulting from the chromosomal translocation t(11;18)(p15;q12) is observed in T-ALL. *SETBP1* (SET binding protein 1, Q9Y6X0) encodes a protein which specifically interacts with SET protein. This latter plays an important role in the regulation of cell death [Panagopoulos et al, 2007]. Another chromosome 18 gene involved in T-ALL is *PTPN2* (P17706) encoding protein

tyrosine phosphatase non-receptor type 2, a cytosolic tyrosine phosphatase that functions as a negative regulator of a variety of tyrosine kinases and other signaling proteins. Deletion of *PTPN2* was specifically found in T-ALL with aberrant expression of the *TLX1* oncogene. Knockdown of *PTPN2* increased the proliferation and cytokine sensitivity of T-ALL cells [Kleppe et al, 2010].

Table 2. Oncogenic translocations involving chromosome 18

Location on the chromosome	Gene or region	Disease	Reference
18q11.2	<i>SYT</i> t(X;18)(p11;q11)	Synovial sarcoma	dos Santos et al, 2001
18q21.1	<i>SETBP1</i> t(11;18)(p15;q12)	T-cell acute lymphoblastic leukemia	Panagopoulos et al, 2007
18q21.3	<i>BCL2</i> t(14;18)(q32;q21)	B-cell lymphoma	Leich et al, 2009

Other somatic genetic events that can occur in cancer tissues include deletions, amplifications and loss of heterozygosity (LOH). It is clear that probable tumor suppressor genes (TSG) undergo deletions and LOH, while oncogene copies may be multiplied by amplification. EPB41L3 protein (erythrocyte membrane protein band 4.1-like 3, Q96HL7) affects tumor cell growth by modulation of post-translational methylation and by increasing of attachment of these cells to a variety of extracellular matrices preventing metastasis. Gene *EPB41L3* (18p.11.3) undergoes allelic losses in **breast cancer** [Kittiniyom et al, 2004]. Another potential tumour suppressor is *L3MBTL4* (human ortholog of lethal (3) malignant brain tumor–like 4 *Drosophila*) on 18p11.31 which is deleted in breast cancer. It is suggested that *L3MBTL4* protein (Q8NA19) interacts with chromatin, possibly

binds methylated histones and thus plays a role in transcriptional regulation of stem cell genes, oncogenes and tumor suppressors [Addou-Klouche et al, 2010] . *DCC* (Deleted in Colorectal Carcinoma, P43146) is the gene of protein which is a receptor with single transmembrane domain. It may function as ligand-dependent suppressor and is often deleted in **colorectal cancer** [Mehlen et al, 2004].

Also deletions are described in different chromosome 18 regions without gene reference, such as frequent LOH on the long arm of chromosome in colorectal cancer [Ogino et al, 2009] and deletions in 18q11.33 for **osteosarcoma** [Johnson-Pais et al, 2003].

There are also a group of potential oncogenes on 18p11.3 for which amplification and overexpression in **esophageal carcinoma** was detected (Nakakuki et al, 2002). Those genes are *YES1* (encoding cellular homolog of the Yamaguchi sarcoma virus oncogene), *HEC* (highly expressed in cancer; plays an important role in chromosome segregation during M phase), *TGIF* (encoding transcription factor modulating the signalling pathway of TGF-beta) and *TYMS* (encoding thymidylate synthase which generates dTMP for DNA synthesis and repair).

The mechanisms of aberrant CpG island methylation in cancer are not fully understood. Sal-like protein 3, also known as zinc finger protein SALL3 (Q9BXA9), is an inhibitory factor for DNA methyltransferase 3 alpha (DNMT3A). SALL3 binds to DNMT3A by a direct interaction and inhibits DNMT3A-mediated CpG island methylation. It was shown that *SALL3* gene (18q23) silenced by associated DNA methylation in **hepatocellular carcinoma (HCC)**. Silencing of *SALL3* results in acceleration of aberrant DNA methylation in HCC [Shikauchi et al, 2009].

Table 3. Somatic mutations on chromosome 18 in cancer

Location on the chromosome	Gene or region	Disease	Reference
18p11.31 18p11.3	<i>L3MBTL4</i> (deletion) <i>EPB41L3</i> (allelic losses)	Breast cancer	Kittiniyom et al, 2004
18p11.3	<i>YES1, TYMS, HEC, TGIF</i> (gene amplification, overexpression)	Esophageal carcinoma	Nakakuki et al, 2002
18q21.1	<i>SMAD4</i> (deletions)	Pancreatic cancer	Blackford et al, 2009
18q21.1 18q21.3 18q	<i>SMAD7</i> (deletions) <i>DCC</i> (deletions) Loss of heterozygosity	Colorectal cancer	Boulay et al, 2003 Mehlen et al, 2004 Ogino et al, 2009
18q21.33	Loss of heterozygosity	Osteosarcoma	Johnson-Pais et al, 2003
18p11.21	<i>PTPN2</i> (deletions)	T-cell acute lymphoblastic leukemia	Kleppe et al, 2010
18q23	<i>SALL3</i> (gene silencing)	Hepatocellular carcinoma	Shikauchi et al, 2009

Several case-control studies revealed associations between SNPs (single nucleotide polymorphisms) in candidate genes of chromosome 18 and cancer development. Such associations were detected for SNPs in *SMAD2* [Slattery et al, 2010a] and *SMAD7* [Slattery et al, 2010b] and **colorectal cancer** and for SNPs

from region nearest to *TCF4* and **prostate cancer** [Murabito et al, 2007]. The question whether these SNPs are driver or passenger mutations is still disputable.

Table 4. Chromosome 18 SNPs associated with cancer

Location on the chromosome	Gene or region	Disease	Reference
18q21	<i>SMAD7</i> (<i>rs12953717</i> <i>rs4939827</i>)	Colorectal cancer	Slattery et al, 2010b
18q21.1	<i>SMAD2</i>	Colorectal cancer	Slattery et al, 2010a
18q21.2	<i>rs 3017183</i> <i>rs 3794889</i> <i>rs 4801149</i> (nearest gene is <i>TCF4</i>)	Prostate cancer	Murabito et al, 2007

As mentioned above, a huge genetic background is described for chromosome 18 in connection with cancer. In scope of the Human Proteome Project, many oncogene and tumor suppressor gene products may be identified and quantified in human specimens to determine their level and study of molecular pathogenesis of particular tumors. A mass spectrometry, such as MRM, has a potential in detecting the cancer-specific protein fusions as a result of oncogenic translocations.

4. NEURODEGENERATIVE AND MENTAL DISORDERS LINKED WITH CHROMOSOME 18.

There are several loci on chromosome 18 that are supposed to be linked with neurodegenerative and mental disorders, such as **Parkinson disease**, **schizophrenia**, **bipolar disorder** and **amyotrophic lateral sclerosis**.

18p11 was repeatedly confirmed as a chromosomal region linked with various hereditary neurological illnesses (see Monogenic hereditary diseases section). The pioneer association and linkage study by Schwab and colleagues (1998) reported that chromosome 18p conferred susceptibility to functional psychoses in families with schizophrenia. As mitochondrial dysfunction was considered to be a risk factor for the onset of schizophrenia and other psychiatric disorders, a core component of mitochondrial respiratory chain, the 24 kD subunit of mitochondrial complex I, NADH-ubiquinone oxidoreductase flavoprotein (NDUFV2, P19404), is a candidate target for psychiatric disorders. *NDUFV2* gene located on 18p11.32-11.31 and contains SNP (rs1156044) in promoter region which is significantly associated with **schizophrenia** [Washizuka et al, 2006] and **bipolar disorder** [Xu et al, 2008]. Another SNP in this gene leading to amino acid change Ala29Val was reported to associate with **Parkinson disease**, degenerative disorder of the central nervous system which impairs motor skills, cognitive processes and behavior functions [Kikuchi et al, 1998].

The second candidate gene for **schizophrenia** in that chromosomal region is *C18orf1* (O15165). *C18orf1* is a novel brain-expressed transcript mapped on 18p11.2. According to its structure, *C18orf1* possibly may bind Ca²⁺ and an unknown ligand. Two SNPs in the gene (-96T/C and 6409 T/C) shown strong association with **schizophrenia** [Kikuchi et al, 2003].

Chromosomal region 18p11.2 is also the linkage region for **psychosis** [Mukherjee et al, 2006] and **bipolar disorder** (locus MAFD1, major affective disorder 1) [Detera-Wadleigh et al, 1999].

Amyotrophic lateral sclerosis (ALS or Lou Gehrig's disease) is a motor neuron disease which causes muscle weakness and atrophy throughout the body caused by degeneration and death of motor neurons. The frequency of ALS is 1-2 cases per 100 000 and most people with ALS die from respiratory failure, usually within 3 to 5 years from the onset of symptoms. Numerous linkage loci for ALS were reported and one of them is located on chromosome 18q21 between markers D18S846 and D18S1109 [Hand et al, 2002].

Table 5. Neurodegenerative and mental disorders linked with chromosome 18

Location on the chromosome	Gene or region	Disease	Reference
18p11.32-11.31	<i>NDUFV2</i>	Parkinson disease Schizophrenia Bipolar disorder	Kikuchi et al, 1998 Washizuka et al, 2006 Xu et al, 2008
18p11.2	<i>C18orf1</i>	Schizophrenia	Kikuchi et al, 2003
18p11.2		Psychosis	Mukherjee et al, 2006
18p	<i>MAFD1</i>	Bipolar disorder	Detera-Wadleigh et al, 1999
18q21	<i>ALS3</i>	Amyotrophic lateral sclerosis	Hand et al, 2002

As for cancer above, the knowledge on genetic aberration in neurodegenerative disorders may be translated to the proteome project.

Proteomics may provide assays to detect disease-related proteins and, in particular, in their mutant forms.

5. OTHER POLYGENIC DISEASE POLYMORPHISMS LINKED WITH THE CHROMOSOME 18.

Polygenic diseases are disorders with strong genetic component which is associated with combined effects of multiple common polymorphisms, each with a small impact on disease risk. Thereby combination of predisposing polymorphisms reflects only higher or lower disease risk rather than obligatory disease development. Most of common disorders associated with chromosome 18 are autoimmune disorders because of several immune-associated genes located on the chromosome. Among those genes are *PTPN2* and *CD226* located on 18p11.21 and 18q22, respectively.

As described above *PTPN2* (tyrosine-protein phosphatase non-receptor type 2, P17706) is a cytosolic tyrosine phosphatase that regulates a variety of cellular processes including cell growth, differentiation, mitotic cycle, and oncogenic transformation. Genome-wide association scans recently identified *PTPN2* as a novel susceptibility gene for several autoimmune disorders. For instance, *PTPN2* associated with **type 1 diabetes** development possibly by modulating IFN- γ signal transduction at the β -cell level [Todd et al, 2007]. Additionally, *PTPN2* regulates cytokine-induced apoptosis and contributes to the pathogenesis of other autoimmune disorders such as **autoimmune thyroiditis (Grave's disease)** [Todd et al, 2007] and **Crohn's disease (inflammatory bowel disease)** [Morgan et al, 2010]. Two SNPs in *PTPN2* gene (rs478582 and rs1893217) show strong association with mentioned diseases.

CD226 (Cluster of Differentiation 226, leukocyte adhesion molecule, DNAM-1, Q15762) is a 65 kDa glycoprotein, member of the immunoglobulin superfamily, expressed on the surface of natural killer cells, platelets, monocytes/macrophages and a subsets of T- and B-cells. CD226 mediates cellular adhesion and interaction of CD226 with its ligands induces NK cell- and CD8+ T cell-mediated cytotoxicity and cytokine secretion. CD226 gene contains SNP (rs763361) leading to amino acid change Ser→ Gly in position 307 of polypeptide chain. It was found association of that SNP with **type 1 diabetes, Grave's disease** [Todd et al, 2007] and **multiple sclerosis** [Hafler et al, 2009].

Also association of SNP (rs2002842) in SALL3 gene (Q9BXA9) with **rheumatoid arthritis** was shown [Julià et al, 2008]. Another autoimmune disorder, **psoriasis**, which characterizes with severe skin inflammation due to accelerated growth cycle and differentiation of skin cells, has been linked with several loci on different chromosomes. One of such loci, *PSORS10* (Psoriasis susceptibility 10), was mapped on 18p11.23 between markers D18S63 and D18S967. It is supposed that *PSORS10* comprised in 3.2 Mb is a minor susceptibility locus for psoriasis [Asumalahti et al, 2003].

RKHD2 (MEX3C) gene on 18q21.1 was identified as linked to **essential hypertension** (Guzman et al, 2006). *RKHD2* (Ring finger and KH domain containing 2, also known as Mex-3 homolog C (*C. elegans*), *MEX3C*, Q5U5Q3) encodes a member of a family of proteins with two K homology (KH) RNA-binding domains and a C-terminal RING-finger domain. The protein interacts with mRNA via the KH domains and shuttles between the nucleus and cytoplasm. In a case-control association study strong association of *MEX3C* haplotype rs1941958G-rs1893379T with essential hypertension was found.

Table 6. Polygenic diseases linked with chromosome 18

Location on the chromosome	Gene or region	Disease	Reference
18p11.21	<i>PTPN2</i>	Type 1 diabetes Grave's disease Crohn's disease	Todd et al, 2007 Todd et al, 2007 Morgan et al, 2010
18q21.2	<i>CD226</i>	Type 1 diabetes Grave's disease Multiple sclerosis	Todd et al, 2007 Todd et al, 2007 Hafler et al, 2009
18q23	<i>SALL3</i>	Rheumatoid arthritis	Julià et al, 2008
18q21.2	<i>RKHD2</i>	Essential hypertension	Guzman et al, 2006
18p11.23	<i>PSORS10</i>	Psoriasis	Asumalahti et al, 2003

6. PROTEINS ENCODED ON THE CHROMOSOME 18 AS DRUG TARGETS²

6.1. Transthyretin.

As described above, transthyretin (TTR, P02766) is a transport protein for thyroxine and retinol. Because of TTR can bind aromatic compounds in the thyroxine binding sites, it can serve as a carrier of numerous small molecules (**pentabromophenol, resveratrol, dibenzofuran-4,6-dicarboxylic acid** and others), non-steroid anti-inflammatory drugs (**diclofenac, diflunisal, flufenamic acid**) and hormones and their synthetic analogs (**diethylstilbestrol, liothyronine, levothyroxine**). TTR can also transport non-aromatic agents such as DMSO

² Mostly adapted from www.drugbank.ca

(dimethyl sulfoxide), beta-mercaptoethanol and acetate ion. A list of drug compounds interacting with TTR is shown in Table 7.

Table 7. Drugs interacting to transthyretin protein.

Drug name	Drug target name
Diethylstilbestrol	Transthyretin
Liothyronine	
Levothyroxine	
Diclofenac	
Diflunisal	
Dimethyl sulfoxide	
3,3',5,5'-Tetraiodothyroacetic Acid	
O-Trifluoromethylphenyl Anthranilic Acid	
Flufenamic Acid	
2,4,6-Tribromophenol	
N-(M-Trifluoromethylphenyl) Phenoxazine-4,6-Dicarboxylic Acid	
Resveratrol	
Beta-Mercaptoethanol	
Pentabromophenol	
3',5'-Dinitro-N-Acetyl-L-Thyronine	
Dibenzofuran-4,6-Dicarboxylic Acid	
Acetate Ion	
2-[3,5-dichloro-4-(2-{2-[2(2-mercaptoethoxy)ethoxy]ethoxy}phenylamino)benzoic acid	

6.2. Proteins involved in cancer

Thymidylate synthase (TYMS, P04818), a key enzyme in *de novo* thymidylate synthesis, converts dUMP to dTMP which is necessary for DNA synthesis and repair. Cytostatic drug methotrexate inhibits several enzymes in folate metabolic pathway, including TYMS, causing defects in DNA and RNA syntheses. Gene TYMS (*TYMS*, 18p11.32) promoter region contains polymorphism 2R/3R, which is 2 or 3 tandem repeats of 28 base pairs. It was observed that

patients with genotype *3R/3R* needed in higher dosage of methotrexate compared with patients carried at least one allele *2R* (Kumagai et al, 2003).

Most of drugs targeted to TYMS are antimetabolites used in anti-cancer chemotherapy. These drugs are **raltitrexed, floxuridine, trifluridine, gemcitabine, pemetrexed and fluorouracil (5-FU)**. **Capecitabine** is chemotherapeutic agent used in the treatment of metastatic breast and colorectal cancers. Capecitabine is a prodrug, that is enzymatically converted to 5-fluorouracil in the tumor. **Folinic acid**, or **leucovorin**, is an adjuvant used in cancer chemotherapy involving the drug methotrexate.

B-cell lymphoma-2 protein described above (encoded by *BCL2* on 18q21.3, P10415) is a target for several chemotherapy anti-mitotic drug (**fludarabine, paclitaxel, docetaxel**) used in the treatment of hematological malignancies. Another one drug interacting with BCL-2, **rasagiline**, is an irreversible inhibitor of monoamine oxidase used as a monotherapy in early Parkinson's disease or as an adjunct therapy in more advanced cases.

Yamaguchi sarcoma virus oncogene homolog (YES1, proto-oncogene tyrosine-protein kinase Yes, P07947) encoded by *YES1* gene on 18p11.3 has tyrosine kinase activity and belongs to the Src family of proteins involved in cancerogenesis. **Dasatinib** is BCR/ABL and Src family tyrosine kinases inhibitor used for chronic myelogenous leukemia (CML) and Philadelphia chromosome-positive acute lymphoblastic leukemia (Ph+ ALL) treatment.

SERPINB5 (serpin peptidase inhibitor, clade B (ovalbumin), member 5, or maspin, P36952). The exact cellular role of SERPINB5 is not currently clear. It was originally described as a breast tumour suppressor, a gene which was active in normal breast epithelial cells and which was down-regulated progressively towards malignancy. It is suggested that maspin suppressed angiogenesis,

reduced tumour growth and metastasis and made cells more sensitive to apoptosis.

Table 8. Drugs interacting to cancer-related proteins encoded by the chromosome 18.

Drug name	Drug target name
Methotrexate	Thymidylate synthase
Raltitrexed	Thymidylate synthase
Floxuridine	Thymidylate synthase
Trifluridine	Thymidylate synthase
Gemcitabine	Thymidylate synthase
Fluorouracil	Thymidylate synthase
Pemetrexed	Thymidylate synthase
Leucovorin	Thymidylate synthase
Capecitabine	Thymidylate synthase
Thymidine-5'-Phosphate	Thymidylate synthase
Ethylene Glycol	Thymidylate synthase
Beta-Mercaptoethanol	Thymidylate synthase
dUMP	Thymidylate synthase
S,S-(2-Hydroxyethyl)Thiocysteine	Thymidylate synthase
Melatonin	Apoptosis regulator Bcl-2
Fludarabine	Apoptosis regulator Bcl-2
Paclitaxel	Apoptosis regulator Bcl-2
Docetaxel	Apoptosis regulator Bcl-2
Rasagiline	Apoptosis regulator Bcl-2
Dasatinib	Proto-oncogene tyrosine-protein kinase Yes
S,S-(2-Hydroxyethyl)Thiocysteine	Serpin B5

6.3. Proteins associated with neurological disorders

Among proteins possibly associated with neurological disorders are two mitochondrial enzymes – AFG3L2 (AFG3-like protein 2, Q9Y4W6) described above and ME-2 (mitochondrial NAD-dependent malic enzyme, P23368), then key enzyme of inositol-signaling pathway, IMPA-2 (inositol monophosphatase 2, O14732) and adenylate cyclase activating polypeptide 1 (ADCYAP1, P18509).

ME-2 is a homotetrameric protein that catalyzes the oxidative decarboxylation of malate to pyruvate. It involves in neuronal synthesis of major inhibitory neurotransmitter gamma-Aminobutyric acid (GABA).

Inositol monophosphatases (IMPases) play a key role in inositol signaling. IMPA-2 defosphorylates myo-inositol in cells. The inositol-signaling pathway is a therapeutic target for **lithium** in the treatment of bipolar disorder because of lithium inhibition of IMPases.

ADCYAP1 stimulates adenylate cyclase and subsequently increases the cAMP level in target cells. ADCYAP1 also functions as a neurotransmitter and neuromodulator. Additionally *ADCYAP1* gene mapped to 18p11.2 in the locus linked with bipolar disorder.

Table 9. Drugs interacting to neurology-related proteins encoded by the chromosome 18.

Drug name	Drug target name
Adenosine triphosphate	AFG3-like protein 2
NADH	NAD-dependent malic enzyme, mitochondrial
Fumarate	NAD-dependent malic enzyme, mitochondrial
Nicotinamide-Adenine-Dinucleotide	NAD-dependent malic enzyme, mitochondrial
Oxalate Ion	NAD-dependent malic enzyme, mitochondrial
Malate Ion	NAD-dependent malic enzyme, mitochondrial
Alpha-Ketomalonic Acid	NAD-dependent malic enzyme, mitochondrial
Tartronate	NAD-dependent malic enzyme, mitochondrial
2,6-Diamino-Hexanoic Acid Amide	Adenylate cyclase-activating polypeptide 1
Lithium	Inositol monophosphatase 2

6.4. Receptors

There are two main receptors encoded by chromosome 18 which can serve as drug targets - MC2R (melanocortin receptor 2, adrenocorticotrophic hormone (ACTH) receptor, Q01718) and HRH4 (hystamine receptor H4, Q9H3N8).

MC2R is a type of melanocortin receptor, which is specific for ACTH. Binding of the receptor by ACTH stimulates the production of cortisol. **Cosyntropin** is a synthetic derivative of ACTH that is used in the ACTH stimulation test to evaluate and diagnose cortisol disorders.

HRH4 is a histamine receptor which is thought to play a role in inflammation and allergy reponses. Activation of this receptor mediates chemotaxis of mast cells and eosinophils and also controls cytokine release from dendritic and T cells. Furthermore, H4 receptors have a role in the differentiation of myeloblasts and promyelocytes. **Clozapine** is an antipsychotic used to treat schizophrenia symptoms in people who have not responded to other medications. Clozapine is a selective monoaminergic antagonist with high affinity for the serotonin Type 2 (5HT₂), dopamine Type 2 (D2), 1 and 2 adrenergic, and H1 histaminergic receptors.

Table 10. Drugs interacting to cell surface receptor proteins encoded by the chromosome 18.

Drug name	Drug target
Cosyntropin	Adrenocorticotrophic hormone receptor
Corticotropin	Adrenocorticotrophic hormone receptor
Clozapine	Histamine H4 receptor

6.5 Proteins associated with cerebrovascular diseases

LMAN1 (lectin, mannose-binding 1, also known as protein ERGIC-53, P49257) mentioned above is a type I integral membrane protein localized in the

intermediate region between the endoplasmic reticulum (ER) and the Golgi complex, recycling between the two compartments for transport FV and FVIII coagulation factors from the ER to the Golgi body.

Tissue plasminogen activator (tPA) is a protein involved in the breakdown of blood clots through catalyze the conversion of plasminogen to plasmin. Recombinant tissue plasminogen activators (r-tPAs) include **alteplase**, **reteplase**, and **tenecteplase** (TNKase) and they are used for treatment of myocardial infarction, embolic and ischemic stroke and other thrombotic disorders. Plasminogen activator inhibitor-2 (placental PAI or serpin peptidase inhibitor, clade B (ovalbumin), member 2, SERPINB2, P05120) is a coagulation factor that inactivates tPA and another plasminogen activator, **urokinase**.

ROCK1 (Rho-associated protein kinase 1, Q13464) is a serine/threonine kinase that contributes to actin stability. Rho-kinase-mediated pathway is mediated enhanced myosin light chain phosphorylation and plays a central role in the pathogenesis of coronary artery spasm. **Hydroxyfasudil**, specific Rho-kinase inhibitor, relaxes arteries mainly by disinhibiting myosin light chain phosphatase through the inhibition of Rho-associated kinase.

Table 11. Drugs interacting to cardiovascular-related proteins encoded by the chromosome 18.

Drug name	Drug target
Antihemophilic Factor	ERGIC-53 protein
Alteplase	Plasminogen activator inhibitor 2
Urokinase	Plasminogen activator inhibitor 2
Reteplase	Plasminogen activator inhibitor 2
Anistreplase	Plasminogen activator inhibitor 2
Tenecteplase	Plasminogen activator inhibitor 2
Hydroxyfasudil	Rho-associated protein kinase 1

6.6. Other drug targets encoded by chromosome 18

Mitochondrial ferrochelatase (FECH, heme synthetase, 18q21.3, P22830) catalyzes the ferrous insertion into protoporphyrin IX. **Cholic acid** is a major primary bile acid produced in the liver. It facilitates fat absorption and cholesterol excretion.

Molybdenum cofactor sulfurase (MOCOS, 18q12, Q96EN8) sulfurates the molybdenum cofactor. Sulfation of molybdenum is essential for xanthine dehydrogenase (XDH) and aldehyde oxidase (ADO) enzymes for their enzymatic activities. MOCOS belongs to the class-V pyridoxal-phosphate-dependent aminotransferase family. **Pyridoxal phosphate** is the active form of vitamin B6 serving as a coenzyme for synthesis of amino acids, neurotransmitters (serotonin, norepinephrine), sphingolipids, aminolevulinic acid.

Cytoplasmic asparaginyl-tRNA synthetase (NARS, 18q21.31, O43776) relates to a class of enzymes that charge tRNAs with their cognate amino acids. Asparaginyl-tRNA synthetase is localized to the cytoplasm and belongs to the class II family of tRNA synthetases.

Cytochrome b5, form A (CYB5A, 18q23, P00167), is a human microsomal cytochrome b5, a membrane bound hemoprotein which function as an electron carrier for several membrane bound oxygenases. Defects in CYB5A are the cause of type IV hereditary methemoglobinemia. It is known that **methoxyflurane** metabolism (*O*-demethylation) by cytochrome P-450 2B4 is markedly stimulated in the presence of CYB5A.

Table 12. Drugs interacting to other proteins encoded by the chromosome 18.

Drug name	Drug target
Cholic Acid	Ferrochelatase, mitochondrial
Pyridoxal Phosphate	Molybdenum cofactor sulfurase
L-Asparagine	Asparaginyl-tRNA synthetase, cytoplasmic
Methoxyflurane	Cytochrome b5

In scope of the proteome project, the quantitative test for drug target proteins encoded in the chromosome 18 may be prepared. Their measurement may be highly useful as a part of personified therapy to predict the patient response to the drug, especially in cancer.

5. ALTERNATIVE SPLICING OF CHROMOSOME 18 PROTEINS IN CANCER.

One of the preferred features of mass-spectrometry-based proteomics is its usability for detection of alternative splice variants in proteins. Peptide sequence-directed approach may be used to screen a significant amount of predicted alternative splicing sites [Mo et al, BMC Bioinformatics, 2008] without need in development of specific affine reagents to each variant [Menon, Omenn, 2011]. The idea of use of alternative splice variants as possible cancer biomarkers was recently suggested and some studies have initiated this direction [Omenn et al, 2010; Menon, Omenn, 2010; Power, McRedmond et al, 2009].

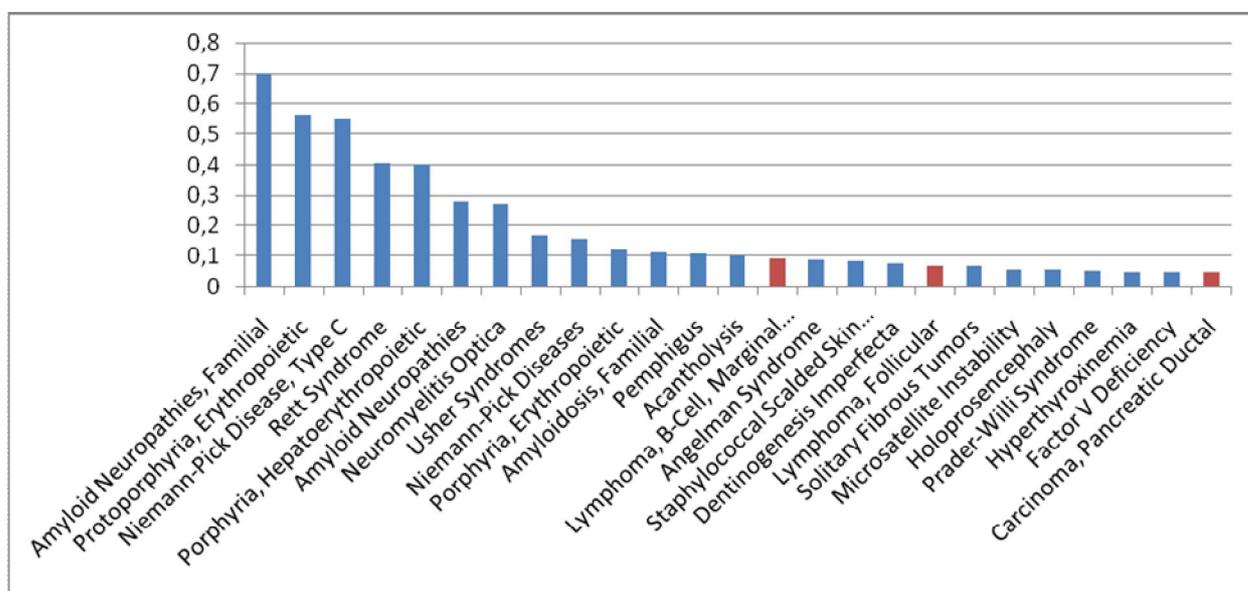
Proteins encoded in chromosome 18 may also be studied from the position of their splice variant expressed. First, exon-exon interfaces for such proteins may be predicted by the described algorithms [Ryan et al, BMC Bioinformatics, 2008; Kim, Lee, Methods Mol Biol, 2008] and signatures of corresponding peptides may be found in mass-spectra of reference and diseased specimens [Tress et al, 2008].

One of the challenges in cancer biomarker field is that most known diagnostic markers change in cancer quantitatively with a many cases in a borderline zone between cancer and control [Gjertson, Albertsen, 2011]. In case of alternative splicing, a chance appears to find a qualitative, highly sensitive diagnostic marker or panel.

6. AUTOMATED TEXT MINING TO SEARCH DISEASE-RELATED PROTEINS ENCODED BY 18TH CHROMOSOME AND ITS COMPARISON TO THE BIOCURATION DATA.

Gene-centric proteome project which is focused on a single chromosome meets a problem of prioritization in terms of diseases that are somehow related to the genes and their products. In order to determine most important disease conditions for the chromosome 18, automated text mining was used. Briefly, the co-occurrence of protein names and disease-related terms were calculated in PubMed paper abstracts. MeSH (Medical Subject Headings) terms were used to search versus protein names as provided by the U.S. National Library of Medicine (<http://www.nlm.nih.gov/pubs/factsheets/mesh.html>).

Fig.1. Top 25 most common MeSH terms that found in PubMed abstracts with chromosome 18 protein names. The y-axis shows a relative number of abstracts with co-occurrence of terms normalized by the overall rate of the term in PubMed abstracts. Cancer-related terms are shown in red, other disease terms are in blue.



As a result, the top 25 of most common MeSH terms related to chromosome 18 proteins was returned. The diagram listing these terms is shown in Fig.1. It can be seen from the figure, that the vast majority of papers describe a link between proteins of interest and amyloidoses, neuropathies and some

hereditary syndromes with a very good correlation to genetic data listed above. Of different type of cancer, only 3 terms in the top 25 are listed and include B-cell lymphoma, follicular lymphoma and pancreatic cancer. Obviously, such an analysis depicts both the biological context and a level of knowledge about diseases.

In order to assess the relevance of automated text mining, its result was compared to the data obtained by expertise. The manual biocuration was done by the expert who used UniProt database annotations for proteins and surfed through the PubMed to annotate each protein of chromosome 18. Diseases found by the biocuration were searched in top 5 of MeSH terms related to each protein, top 15 of such terms and the rest of terms, respectively. About 40% of manually found terms occur in the top 5 list, 73% of such terms were at least in top 15, whereas 14% of manually found data were skipped by automated analysis. The level of correspondence between 2 strategies of annotation may be considered as satisfactory.

In summary, automated data mining found that, of socially important diseases, the chromosome 18 proteins are mostly important in various neuropathies and lymphomas.

The Addendum working group

S.Moshkovskii, O.Voronko, E.Ponomarenko, M.Pyatnitsky, M.Karpova, E.Ilgisonis, A.Archakov

Version 5, May. 6, 2011