Journal of Experimental Therapeutics and Oncology, Vol. 8, pp. 65–76
Reprints available directly from the publisher
Photocopying permitted by license only

SHORT COMMUNICATION

Quantifying the possible cross-reactivity risk of an HPV16 vaccine

Darja Kanduc

Department of Biochemistry and Molecular Biology, University of Bari, Italy

Correspondence to: D. Kanduc, Department Biochemistry and Molecular Biology, University of Bari, Bari 70126 – Italy. Telephone: 39 080 544 3321; Fax: 39 080 544 3321. Email: d.kanduc@biologia.uniba.it; dkanduc@gmail.com

(Received: March 18, 2009; accepted: May 5, 2009)

Background: The potential adverse events associated with vaccination for infectious diseases underscore the need for effective analysis and definition of possible vaccine side effects. Using the HPV16 proteome as a model, we quantified the actual and theoretical risks of anti-HPV16 vaccination, and defined the potential disease spectrum derived from concomitant cross-reactions with the human organism.

Methods: We searched the primary sequence of the HPV16 proteome for heptamer aminoacid sequences shared with human proteins using the Protein International Resource database.

Results: The human proteome contains 82 heptapeptides and two octapeptides found in HPV16. The viral matches are spread among proteins involved in fundamental processes, such as cell differentiation and growth and neurosensory regulation. The human proteins containing the HPV16-derived heptamers include cell-adhesion molecules, leukocyte differentiation antigens, enzymes, proteins associated with spermatogenesis, transcription factors, and neuronal antigens. The number of viral matches and their locations make the occurrence of side autoimmune cross-reactions in the human host following HPV16-based vaccination almost unavoidable.

Conclusions: Any antigen-based vaccine needs to be carefully and thoroughly designed and critically screened for potential side effects by comparing sequence similarity at the molecular level.

Key words: HPV16 proteome; human proteome; similarity analysis; viral versus human proteome overlapping; vaccine-related cross-reactions

Vaccination for infectious diseases is associated with potential adverse events.[1] Indeed, current antigen-specific immunotherapy protocols may target not only the antigen from the infectious microorganism, but also host tissues expressing antigens that share sequences with the target.[2] There is no clear, defined, mathematical tabulation of the possible cross-reactivity risks associated with a vaccination protocol. This task is not unachievable, because the entire protein assemblies of viruses, bacteria, and higher vertebrates have been sequenced, allowing for proteomic sequence-to-sequence profiling analyses.

Here, the HPV16 polyprotein was examined for amino acid sequence similarity to the human proteome at the heptamer level. A high level of sharing of heptapeptide motifs between HPV16 and human proteins was found. This is discussed in relationship to the potential cross-reactivity risk of HPV16 vaccine.

METHODS

The HPV16 oncoprotein primary sequence (Medline accession no. K02718) was dissected into heptamers that were analyzed for sequence similarity to the human proteome using PIR perfect match (pir.georgetown.edu/pirwww/search/fasta). The heptamers were offset by one residue, i.e., overlapping by six residues: i.e., MHQKRTA HQKRTAM, QKRTAMF, KRTAMFQ,

etc. The human proteome was obtained from the EBI Integr8 site (http://www.ebi.ac.uk/integr8) and consisted of 34,044 non-redundant proteins at the time of download. The function of the human proteins and potential disease associations were analyzed using the Universal Protein Resource (UniProt; www.uniprot.org/uniprot).

RESULTS AND DISCUSSION

Using the HPV16 proteome as a model, we quantified the potential cross-reactivity risk resulting from anti-viral vaccination. HPV16 proteins were analyzed for amino acid sequence identity to the human proteome using heptamers as scanning units.

Table 1 lists the human proteins containing exact HPV16 heptamer matches. Several human proteins share identical heptapeptide motifs with HPV16 polyprotein. Although the theoretical probability of 20 amino acids occurring in seven identical residues between two proteins is 1 in 20^7 = 1,280,000,000, the viral L1, L2, E6, and E7 proteins actually share heptapeptide sequences in common with the human proteome for a total of 82 perfect overlaps. Moreover, although the theoretical probability of 20 amino acids occurring in eight identical residues between two proteins is 1 in 20^8 = 25,600,000,000, two viral octamers are present in the human proteome: (1) The HPV16 $L2_{67\text{-}74}$ GGRTGYIP peptide occurs in the peripheral benzodiazepine receptor-interacting protein, which is highly expressed in the temporal lobe and putamen;[19] and (2) the HPV16 $E7_{85\text{-}92}$ DGLTVIVT peptide sequence occurs in γ-glutamyltransferase 7, a marker of hepatic carcinogenesis.[70] The discrepancy between the actual and theoretical values for heptapeptide overlap between the viral and human proteomes is high. This discrepancy underscores the fact that a higher potential risk level is expected. Based on the need for five or six amino acids to induce a monoclonal antibody response,[2] the 82 heptapeptide overlaps can clearly induce autoimmune reactions.

Likewise, the potential pathological burden is heavy. Table 1 shows that viral matches are present in human proteins involved in cell growth and differentiation, spermatogenesis, developing and regenerating muscles, the morphogenetic programming of tissues, lipoprotein metabolism, the early stage of adipogenesis, and survival motor neuron protein. Moreover, alterations of the human proteins hosting the HPV16 motifs produce a number of pathologies that range from muscle diseases to neurological disturbances. Consequently, it is logical to postulate that targeting these human antigens might induce many of the syndromes listed in Table 1. As detailed in Table 1, autoimmune reactions against the human proteins hosting HPV16-derived motifs might lead to pathologies including spinal muscular atrophy, proximal muscle weakness causing a waddling gait, toe-walking, lordosis, frequent falls, difficulty standing up and climbing stairs, cardiovascular and musculoskeletal abnormalities, disorders of lipoprotein metabolism leading to hypercholesterolemia, and increased proneness to coronary artery disease.

Table 1. Sharing of 7-mer motifs between HPV16 and human proteomes.

Heptapeptide motifs are described by their location in the viral protein, amino acid sequence, and number of occurrences in the human proteome. Description of the human proteins hosting heptapeptides from HPV16 proteome comprehends accession number, (potential) function, cellular location, available data on disease association (from UniprotKB, www.uniprot.org/uniprot).

Viral Protein	Aa Pos	Sequence	N°	Human proteins hosting eptapeptides from HPV16 proteome	Ref.
L1	41	PVPVSKV	1	EPS8_HUMAN: Epidermal growth factor receptor kinase substrate 8. Cell proliferation: upon binding to EGF receptor enhances EGF-dependent mitogenic signals. Expressed in all tissues analyzed.	3
L1	63	AGTSRLL	1	Q9C0C1: Protein dispatched homolog 2. Multi-pass membrane protein.	4
L1	238	TLQANKS	1	P25391 - LAMA1_HUMAN: Laminin subunit α -1. Laminin is a complex glycoprotein, consisting of three different polypeptide chains (α, β, γ). α-1 is a subunit of laminin-1 (EHS laminin) and laminin-3 (S-laminin). Laminin mediates the attachment and organization of cells into tissues during embryonic development. Regulation of epithelial cell proliferation, cell adhesion, cell migration, embryonic development.	5

Viral Protein	Aa Pos	Sequence	N°	Human proteins hosting eptapeptides from HPV16 proteome	Ref.
L1	243	KSEVPLD	1	ELMO3_HUMAN: Engulfment and cell motility protein 3. Involved in cytoske-letal rearrangements during phagocytosis of apoptotic cells and cell motility.	6
L1	292	AVGENVP	1	BACH2_HUMAN: Transcription regulator protein BACH2. DNA-dependent.	7
L1	304	KGSGSTA	1	Q9ULG1: Putative DNA helicase component of the INO80 complex which remodels chromatin by shifting nucleosomes. Recruited by YY1 to YY1-activated genes, where it acts as an essential coactivator. Binds DNA. Nucleus.	8
L1	305	GSGSTAN	1	RCBT1_HUMAN: Regulator of chromosome condensation and BTB domain-containing protein 1. Chronic lymphocytic leukemia deletion region gene 7 protein. Involved in cell cycle regulation by chromatin remodelling. Ubiquitously expressed.	9
L1	319	PTPSGSM	1	ANR53_HUMAN: Ankyrin repeat domain-containing protein 53.	6
L1	354	QLFVTVV	1	ANR17_HUMAN: Ankyrin repeat domain-containing protein 17. Gene trap ankyrin repeat protein. Serologically defined breast cancer antigen NY-BR-16. Earliest specific in situ marker of hepatic differentiation during embryogenesis, useful for characterization of inductive events involved in hepatic specification. Target of enterovirus 71 which is the major etiological agent of HFMD (hand, foot and mouth disease). Cytoplasm. Nucleus.	10
L1	372	AAISTSE	1	ANR11_HUMAN: Ankyrin repeat domain-containing protein 11. May recruit HDACs to the p160 coactivators/nuclear receptor complex to inhibit ligand-dependent transactivation. Nucleus.	11
L1	398	LQFIFQL	1	P04114 – APOB_HUMAN: Apolipoprotein B-100. Apolipo-protein B is a major protein constituent of chylomicrons (apo B-48), LDL (apo B-100) and VLDL (apo B-100). Apo B-100 functions as a recognition signal for the cellular binding and internalization of LDL particles by the apoB/E receptor. Defects in APOB are a cause of familial hypobetalipoproteinemia (FHBL). Subjects have extremely low plasma LDL cholesterol and apoB-100 concentrations, and clinical presentation may vary from no symptoms to severe gastrointestinal and neurological dysfunction similar to abetalipoproteinemia. Defects in APOB are a cause of familial ligand-defective apolipoprotein B-100 (FDB). FDB is a dominantly inherited disorder of lipoprotein metabolism leading to hypercholesterolemia and increased proneness to coronary artery disease (CAD). The plasma cholesterol levels are dramatically elevated due to impaired clearance of LDL particles by defective APOB/E receptors.	12
L1	432	LQPPPGG	2	PBX4_HUMAN: Pre-B-cell leukemia transcription factor 4. Q8IXF5: Sharpin. Shank-associated RH domain-interacting protein. Has a role in immune development and control of inflammation. Expressed in skeletal muscle and placenta and at lower levels in brain, heart, colon without mucosa, thymus, spleen, kidney, liver, small intestine, lung and peripheral blood leukocytes.	13 14
L1	436	PGGTLED	1	Q96SF2: Putative T-complex protein 1 subunit θ-like 2. Poss-ible molecular chaperone; assists the folding of proteins upon ATP hydrolysis.	15

Table 1. *Continued.*

Viral Protein	Aa Pos	Sequence	N°	Human proteins hosting eptapeptides from HPV16 proteome	Ref.
L1	479	EKFSADL	1	SPT6H_HUMAN: Transcription elongation factor SPT6. Tat-cotransactivator 2 protein. Acts to stimulate transcriptional elongation by RNA polymerase II.	16
L1	512	KATPTTS	1	C1QR1_HUMAN: Complement component C1q receptor. Receptor for C1q, mannose-binding lectin and pulmonary surfactant protein A. May mediate the enhancement of phagocytosis in monocytes and macrophages upon interaction with soluble defense collagens. Highly expressed in endothelial cells, platelets, monocytes and neutrophils.	17
L1	515	PTTSSTS	1	Q59EI7: Cytoplasmic nuclear factor of activated T-cells 3 isoform 4 variant. Regulation of transcription, DNA-dependent.	18
L2	67	GGRTGYI	1	RIMB1_HUMAN: Peripheral benzodiazepine receptor-interacting protein. Cytoplasm. Mitochondrion. Expressed in brain, pituitary gland and thymus in adult. In brain adult, high expression found in temporal lobe and the putamen, followed by amygdala, caudate nucleus, cerebral cortex, occipital and frontal lobe. High expression level in fetal tissues like brain, heart, kidney and thymus.	19
L2	68	GRTGYIP	1	RIMB1_HUMAN: see above.	19
	99	GPSDPSI	1	CDYL1_HUMAN: Repressor of transcription: histone acetyltransferase activity. Spermatogenesis. Ubiquitous. Expressed at moderate levels in all tissues examined.	20
L2	105	IVSLVEE	1	Q53R12: Transmembrane 4 L6 family member 20.	21
L2	106	VSLVEET	1	SYNE2_HUMAN: Nuclear envelope spectrin repeat protein 2. Synaptic nuclear envelope protein 2. Involved in the maintenance of nuclear organization and structural integrity. Probable anchoring protein which theters the nucleus to the cytoskeleton. Connects nuclei to the cytoskeleton by interacting with the nuclear envelope and with F-actin in the cytoplasm. Remains associated with the nuclear envelope during its breakdown in mitotic cells. Widely expressed, with higher level in kidney, adult and fetal liver, stomach and placenta.	22
L2	128	PDVSGFS	1	Q6ZSA9: 3-hydroxymethyl-3-methylglutaryl-CoA lyase-like protein 1.	6
L2	150	NTVTTVT	1	O15026: Helicase SRCAP. Catalytic component of the SRCAP complex which mediates the ATP-dependent exchange of histone H2AZ/H2B dimers for nucleosomal H2A/H2B, leading to transcriptional regulation of selected genes by chromatin remodeling. Acts as a coactivator for CREB-mediated transcription, steroid receptor-mediated transcription, and Notch-mediated transcription.	23
L2	151	TVTTVTT	2	DMD_HUMAN: Dystrophin. Plays a role in anchoring the cytoskeleton to the plasma membrane. Defects in dystrophin are the cause of 1) Duchenne muscular dystrophy (DMD). It typically presents as proximal muscle weakness causing waddling gait, toe-walking, lordosis, frequent falls, and difficulty in standing up and climbing up stairs. The pelvic girdle is affected first, then the shoulder girdle. Progression is steady: most patients are confined to a wheelchair by age of 10 or 12. Contractures and scoliosis. 2) Becker muscular dystrophy (BMD). BMD resembles DMD in hereditary and clinical features but is later in onset and more benign. 3) cardiomyopathy dilated X-linked	24

Viral Protein	Aa Pos	Sequence	N°	Human proteins hosting eptapeptides from HPV16 proteome	Ref.
				type 3B; also known as X-linked dilated cardiomyopathy. Dilated cardiomyopathy is characterized by ventricular dilation and impaired systolic function, resulting in congestive heart failure and arrhythmia. Patients are at risk of premature death. Q7L0R7: RING finger protein 44. Zinc ion binding protein.	 25
L2	164	TDPSVLQ	2	DC1I1_HUMAN: Cytoplasmic dynein 1 intermediate chain 1. The intermediate chains seem to help dynein bind to dynactin 150 kDa component. May play a role in mediating the interaction of cytoplasmic dynein with membranous organelles and kinetochores. DC1I2_HUMAN: Cytoplasmic dynein 1 intermediate chain 2.	26 26
L2	232	TQQVKVV	1	Q6PIS3: XYLB protein. Xylulose kinase. Carbohydrate metabolic process.	27
L2	355	AASPTSI	1	NEO1_HUMAN: Neogenin. Regulatory protein in the transition of undifferentiated proliferating cells to their differentiated state. Functions as a cell adhesion molecule in a broad spectrum of embryonic and adult tissues. Widely expressed.	28
L2	382	VPSVPST	1	Q5VT65: Pleckstrin homology domain-containing family M member 2. Salmonella-induced filaments A and kinesin-interacting protein.	29
L2	406	YNIPLVS	1	SNRK_HUMAN: SNF-related serine/threonine-protein kinase. Plays a role in hematopoietic cell proliferation or differentiation. Potential mediator of neuronal apoptosis. Expressed in hematopoietic progenitor cells and leukemic cells.	30
L2	455	LRKRRKR	1	BTC_HUMAN: Probetacellulin. Potent mitogen for retinal pigment epithelial cells and vascular smooth muscle cells. The effects are probably mediated by the EGF receptor and other related receptors. Expressed in pancreas and small intestine.	31
L2	467	SDVSLAA	1	Q29SS9: Nck-associated protein 5. Peripheral clock protein. Expressed in fetal and adult brain, leukocytes and fetal fibroblasts.	32
E1	21	VEAVVEK	1	PNKP_HUMAN: Bifunctional polynucleotide phosphatase/kinase. Function in DNA repair following ionizing radiation or oxidative damage. Expressed in many tissues with highest expression in spleen, testis, pancreas, heart and kidney.	33
E1	48	LVDFIVN	1	Q5H9M0: MUM1-like protein 1: Mutated melanoma-associated antigen 1-like protein 1. Phosphorylated upon DNA damage, probably by ATM or ATR.	6
E1	129	SEDSGYG	1	GT2D1_HUMAN: General transcription factor II-I repeat domain-containing protein 1. Slow-muscle-fiber enhancer-binding protein. Williams-Beuren syndrome chromosomal region 11 protein. Involved in cell-cycle progression and skeletal muscle differentiation. May contribute to slow-twitch fiber type specificity during myogenesis and in regenerating muscles. Binds troponin I slow-muscle fiber enhancer. Binds specifically and with high affinity to the EFG sequences derived from the early enhancer of HOXC8. Interacts with the retinoblastoma protein via its C-terminus. Highly expressed in adult skeletal muscle, heart, fibroblast, bone and fetal tissues. Highly expressed in developing and regenerating muscles, at	34 47-50

Table 1. *Continued.*

Viral Protein	Aa Pos	Sequence	N°	Human proteins hosting eptapeptides from HPV16 proteome	Ref.
				the time of myofiber diversification. Phosphorylated upon DNA damage, probably by ATM or ATR. Haploinsufficiency of GTF2IRD1 may be the cause of cardiovascular and musculoskeletal abnormalities observed in Williams-Beuren syndrome (WBS), a rare developmental disorder.	
E1	137	TEVETQQ	1	Q86W99: 5-azacytidine-induced protein 2 (AZI2). NF-κ-B-activating kinase-associated protein 1. Activates serine/threonine-protein kinase TBK1 and facilitates its oligomerization. Enhances the phosphorylation of NF-κ-B p65 subunit RELA by TBK1. Promotes TBK1-induced as well as TNF-α or PMA-induced activation of NF-κ-B. Participates in IFN- β promoter activation via TICAM1. Widely expressed. Abundant expression seen in the pancreas and testis. Cytoplasm.	35
E1	161	SGGSGGG	16	CUX1_HUMAN: Homeobox protein cut-like 1. CCAAT displacement protein. Potential broad role in mammalian development as a repressor of developmentally regulated gene expression. May act by preventing binding of positively-activing CCAAT factors to promoters. Component of nf-munr repressor; binds to the matrix attachment regions (MARs) (5' and 3') of the immunoglobulin heavy chain enhancer. Represses T-cell receptor (TCR) β enhancer function by binding to MARβ, an ATC-rich DNA sequence located upstream of the TCR β enhancer.	36
				DNJB1_HUMAN: Heat shock 40 kDa protein 1. Stress response.	37
				HXB3_HUMAN: Homeobox protein Hox-B3. Sequence-specific transcription factor which is part of a developmental regulatory system that provides cells with specific positional identities on the anterior-posterior axis. Expressed in whole embryos and fetuses at 5-9 weeks from conception.	38
				K1C9_HUMAN: Keratin, type I cytoskeletal 9. Important function either in the mature palmar and plantar skin tissue or in the morphogenetic program of the formation of these tissues. Defects in KRT9 are a cause of palmoplantar keratoderma epidermolytic (EPPK), a dermatological disorder characterized by diffuse thickening of the epidermis on the entire surface of palms and soles sharply bordered with erythematous margins.	39
				KCNA4_HUMAN: Potassium voltage-gated channel subfamily A member 4. Mediates the voltage-dependent potassium ion permeability of excitable membranes.	40
				MAST3_HUMAN: Microtubule-associated Ser/Thr-protein kinase 3. I nteracts with PTEN.	41
				Q53EP0: Fibronectin type III domain-containing protein 3B. Factor for adipocyte differentiation 104. Predominantly expressed in white adipose tissue (WAT). Expression increased in the early stage of adipogenesis.	42
				Q71RC4: PP13187. With function of inhibiting cancer cell growth.	43

Viral Protein	Aa Pos	Sequence	N°	Human proteins hosting eptapeptides from HPV16 proteome	Ref.
				Q8N849: cDNA FLJ40018 fis, clone STOMA2006398	6
				Q8TB79: CXXC-type zinc finger protein 5. Putative NF-κ-B-activating protein 102. Putative MAPK-activating protein PM08. May indirectly participate in activation of the NF-κ-B and MAPK pathways.	44
				Q96JN8: NHR domain-containing protein KIAA1787. Widely expressed at high levels.	45
				Q96PX6: Coiled-coil domain-containing protein 85°.	46
				SMN_HUMAN: Survival motor neuron protein. The SMN complex plays an essential role in spliceosomal snRNP assembly in the cytoplasm and is required for pre-mRNA splicing in the nucleus. Highly expressed in brain, spinal cord, kidney, liver; moderate levels in skeletal and cardiac muscle; low levels in fibroblasts and lymphocytes. Defects in SMN1 are the cause of spinal muscular atrophies (SMA). SMAs are fatal autosomal recessive disorders subclassified as type 1 (SMA1, Werdnig-Hoffmann disease), type 2 (SMA2, intermediate form), type 3 (SMA3, Wohlfart-Kugelberg-Welander disease) and type 4 (SMA4, adult form) based upon the age of onset and clinical severity. These neurodegenerative disorders are characterized by degeneration of lower motor neurons, leading to progressive paralysis and muscular atrophy.	47
				SP8_HUMAN: Transcription factor Sp8. Specificity protein 8.	48
				SYGP1_HUMAN: Synaptic Ras GTPase-activating protein 1. Neuronal RasGAP. Major constituent of the PSD essential for postsynaptic signaling. Inhibitory regulator of the Ras-cAMP pathway. Member of the NMDAR signaling complex in excitatory synapses, it may play a role in NMDAR-dependent control of AMPAR potentiation, AMPAR membrane trafficking and synaptic plasticity. Regulates AMPAR-mediated miniature excitatory postsynaptic currents. May be involved in certain forms of brain injury, leading to long-term learning and memory deficits.	45
				TRIO_HUMAN: Triple functional domain protein. PTPRF-interacting protein. Promotes the exchange of GDP by GTP. Together with leukocyte antigen-related (LAR) protein, it could play a role in coordinating cell-matrix and cytoskeletal rearrangements.	49
E1	189	TPLTNIL	1	Q685J3 - MUC17_HUMAN: Mucin-17. Small intestinal mucin-3. Probably plays a role in maintaining homeostasis on mucosal surfaces. Expressed almost exclusively in the intestine. Expression is especially high in both the duodenum and transverse colon. Expressed in mature absorptive cells of the small intestinal villi. No expression in goblet cells. Highly expressed in pancreatic adenocarcinoma tissue (at protein level). Not detected in normal pancreas, in pancreatitis or in cell lines derived from other cancers.	50
E1	193	NILNVLK	1	CCNB3_HUMAN: G2/mitotic-specific cyclin-B3. Cyclins play an essential role in the control of the cell cycle, notably via their destruction during cell division. Its tissue specificity suggest that it may be required during early meiotic prophase I. Testis specific. In testis, it is expressed in developing germ cells, but not in Leydig cells.	51

Table 1. *Continued.*

Viral Protein	Aa Pos	Sequence	N°	Human proteins hosting eptapeptides from HPV16 proteome	Ref.
E1	209	AKFKELY	1	RABE1_HUMAN: Rab GTPase-binding effector protein 1. Renal carcinoma antigen NY-REN-17. Involved in endocytic membrane fusion and membrane trafficking of recycling endosomes.	52
E1	223	VRPFKSN	1	CX023_HUMAN: Uncharacterized protein CXorf23. Mitochondrial component.	53
E1	289	TIEKLLS	1	Q86TI0 - TBCD1_HUMAN: TBC1 domain family member 1.	54
E1	290	IEKLLSK	1	SOX5_HUMAN: Transcription factor SOX-5. Binds specifically to the DNA sequence 5'-AACAAT-3'. Activates transcription of COL2A1 and AGC1 in vitro.	55
E1	310	PKLRSTA	1	WIPF1_HUMAN: Wiskott-Aldrich syndrome protein-interacting protein. Activity on the actin cytoskeleton. Induces actin polymerization and redistribution. May participate in regulating the subcellular localization of WASL, resulting in the disassembly of stress fibers in favor of filopodia formation. Plays an important role in the intracellular motility of vaccinia virus by functioning as an adapter for recruiting WASL to vaccinia virus. Colocalized with actin stress fibers. When co-expressed with WASL, no longer associated with actin filaments but accumulated in perinuclear and cortical areas like WASL. Highly expressed in peripheral blood mononuclear cells, spleen, placenta, small intestine, colon, thymus. Lower expression in ovary, heart, brain, lung, liver, skeletal muscle, kidney, pancreas, prostate and testis.	56
E1	459	ALKRFLQ	1	Q86UQ4 -ABCAD_HUMAN: ATP-binding cassette sub-family A member 13. Multi-pass membrane protein. Expressed in testis, bone marrow and trachea.	57
E1	609	WKSFFSR	1	Q9P2K1: Coiled-coil and C2 domain-containing protein. Strongly expressed in prostate, pancreas, kidney, lung and liver. Defects in CC2D2A are the cause of Meckel syndrome type 6 (MKS6), an autosomal recessive disorder characterized by renal cysts and variably associated features including developmental anomalies of the central nervous system (encephalocele), hepatic ductal dysplasia and cysts, and polydactyly. Defects in CC2D2A are the cause of mental retardation (MR) and retinitis pigmentosa (RP). Non-syndromic MR is characterized by significantly sub-average general intellectual functioning associated with impairments in adaptative behavior and manifested during the developmental period. RP leads to degeneration of retinal photoreceptor cells. Patients have night vision blindness and loss of midperipheral visual field. Progressively they lose their far peripheral visual field and eventually central vision.	58
E1	620	LSLHEDE	1	GALC_HUMAN: Galactocerebrosidase: responsible for the lysosomal catabolism of galactosylceramide, a major lipid in myelin, kidney, epithelial cells of small intestine and colon. Hydrolyzes the galactose ester bonds of galactosylceramide, lactosylceramide, galactosylsphingosine. Defects in GALC are the cause of globoid cell leukodystrophy or Krabbe disease. This disorder results in the insufficient catabolism of several galactolipids that are important in the production of normal myelin. Clinically, the most frequent form is the infantile form. Most patients (90%) present before 6 months of age with	59

Viral Protein	Aa Pos	Sequence	N°	Human proteins hosting eptapeptides from HPV16 proteome	Ref.
				irritability, spasticity, arrest of motor and mental development, and bouts of temperature elevation without infection. This is followed by myoclonic jerks of arms and legs, oposthotonus, hypertonic fits, and mental regression, which progresses to a severe decerebrate condition with no voluntary movements and death from respiratory infections or cerebral hyperpyrexia before 2 yrs of age. Cases with later onset, presenting with unexplained blindness, weakness and/or progressive motor, and sensory neuropathy that can progress to severe mental incapacity and death, have been identified.	
E1	626	EDKENDG	1	MPIP2_HUMAN: M-phase inducer phosphatase 2. Tyrosine protein phosphatase which functions as a dosage-dependent inducer of mitotic progression. Dephosphorylates CDC2 and stimulates its kinase activity. Cell division.	60
E2	61	TLAVSKN	1	COG5_HUMAN: Conserved oligomeric Golgi complex subunit 5. Component required for normal Golgi morphology and localization.	61
E2	111	KKHGYTV	1	S26A2_HUMAN: Sulfate transporter. Diastrophic dysplasia protein. May play a role in endochondral bone formation. Ubiquitously expressed. Defects in SLC26A2 are the cause of: 1) diastrophic dysplasia, an autosomal recessive disease characterized by osteochondrodysplasia with clinical features including dwarfism, spinal deformation, and specific joint abnormalities; 2) achondrogenesis type 1B, a recessively inherited chondrodysplasia characterized by extremely poor skeletal development and perinatal death; 3) atelosteogenesis type 2 (also known as neonatal osseous dysplasia 1), characterized by severely shortened limbs, small chest, scoliosis, club foot of the equinovarus type, abducted thumbs and great toes, and cleft palate. Patients die of respiratory insufficiency shortly after birth because of the collapse of the airways and pulmonary hypoplasia due to the small rib cage; 4) multiple epiphyseal dysplasia type 4, a generalized skeletal dysplasia associated with significant morbidity. Joint pain and deformity, waddling gait, and short stature are the main clinical signs and symptoms. Is categorized into the more severe Fairbank and the milder Ribbing types.	62
E2	259	RDSVDSA	1	TPRGL_HUMAN: Tumor protein p63-regulated gene 1-like protein. Localized to presynaptic nerve terminals.	63
E2	330	SAIVTLT	1	P210L_HUMAN: Nuclear pore membrane glycoprotein 210-like.	64
E6	101	KPLCDLL	3	IMA1_HUMAN: Importin subunit α-1. Functions in nuclear protein import as an adapter protein for nuclear receptor KPNB1. IMA5_HUMAN - Importin subunit α-6. Details as for IMA1_HUMAN. IMA7_HUMAN – Importin subunit α-7. Details as for IMA1_HUMAN.	65
E7	27	QLNDSSE	2	NPT4_HUMAN – Sodium-dependent phosphate transport protein 4. Expressed mainly in the liver and kidney, very low levels in the small intestine and testis. WDR67_HUMAN: WD repeat-containing protein 67. Regulation of Rab GTPase.	66 27

Table 1. *Continued.*

Viral Protein	Aa Pos	Sequence	N°	Human proteins hosting eptapeptides from HPV16 proteome	Ref.
E7	28	LNDSSEE	1	PRDM2_HUMAN: Retinoblastoma protein-interacting zinc finger protein S-adenosyl-L-methionine-dependent histone methyltransferase that specifically methylates Lys-9 of histone H3. Binds to the macrophage-specific TPA-responsive element (MTE) of the HMOX1 (heme oxygenase 1) gene and acts as a transcriptional activator of this gene.	67
E7	30	DSSEEED	1	Q9NXF7: Uncharacterized protein C4orf30.	6
E7	31	SSEEEDE	2	ARI4A_HUMAN: AT-rich interactive domain-containing protein 4A. Retinoblastoma-binding protein 1. Transcription regulation. ERCC5_HUMAN: DNA repair protein complementing XP-G cells. Cofactor for a DNA glycosylase that removes oxidized pyrimidines from DNA. Defects in ERCC5 cause xeroderma pigmentosum complementation group G, characterized by solar hypersensitivity of the skin, predisposition for developing cancers on areas exposed to sunlight and neurological abnormalities. Patients may present features of Cockayne syndrome, dwarfism, pigmentary retinopathy, sensorineural deafness, microcephaly, mental retardation, ataxia, decreased nerve conduction.	68 69
E7	28	PRPIPKP	1	Q96MJ9: cDNA FLJ32252 fis, clone PROST1000167, similar to SYNAPSIN I.	12
E7	85	DGLTVIV	1	GGT7_HUMAN: γ-glutamyltransferase 7.	70
E7	86	GLTVIVT	1	GGT7_HUMAN: γ-glutamyltransferase 7.	70
E7	87	LTVIVTL*	1	Q13973: Leukocyte differentiation antigen. CD1e antigen.	71

*Present in all isoforms.

REFERENCES

1. www.cdc.gov/ vaccinesafety/
2. Oldstone MB. Molecular mimicry and immune-mediated diseases. Faseb J 1998;12:1255–65.
3. Wong WT, Carlomagno F, Druck T, et al. Evolutionary conservation of the EPS8 gene and its mapping to human chromosome 12q23-q24. Oncogene 1994;9:3057–61.
4. Nagase T, Kikuno R, Hattori A, Kondo Y, Okumura K, Ohara O. Prediction of the coding sequences of unidentified human genes. XIX. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro. DNA Res 2000;7:347–55.
5. Olsen D, Nagayoshi T, Fazio M, et al. Human laminin: cloning and sequence analysis of cDNAs encoding A, B1 and B2 chains, and expression of the corresponding genes in human skin and cultured cells. Lab Invest 1989;60:772–82.
6. Ota T, Suzuki Y, Nishikawa T, et al. Complete sequencing and characterization of 21,243 full-length human cDNAs. Nat Genet 2004;36:40–5.
7. Vieira SA, Deininger MW, Sorour A, et al. Transcription factor BACH2 is transcriptionally regulated by the BCR/ABL oncogene. Genes Chromosomes Cancer 2001;32:353–63.
8. Jin J, Cai Y, Yao T, et al. A mammalian chromatin remodeling complex with similarities to the yeast INO80 complex. J Biol Chem 2005;280:41207–12.
9. Mabuchi H, Fujii H, Calin G, et al. Cloning and characterization of CLLD6, CLLD7, and CLLD8, novel candidate genes for leukemogenesis at chromosome 13q14, a region commonly deleted in B-cell chronic lymphocytic leukemia. Cancer Res 2001;61:2870–7.
10. Scanlan MJ, Gout I, Gordon CM, et al. Humoral immunity to human breast cancer: antigen definition and quantitative analysis of mRNA expression. Cancer Immun 2001;1:4.
11. Zhang A, Yeung PL, Li CW, et al. Identification of a novel family of ankyrin repeats-containing cofactors for p160 nuclear receptor coactivators. J Biol Chem 2004;279:33799–805.
12. Chen SH, Yang CY, Chen PF, et al. The complete cDNA and amino acid sequence of human apolipoprotein B-100. J Biol Chem 1986;261:12918–21.

13. Wagner K, Mincheva A, Korn B, Lichter P, Poepperl H. Pbx4, a new Pbx family member on mouse chromosome 8 is expressed during spermatogenesis. Mech Dev 2001;103:127–31.

14. Daigo Y, Takayama I, Ward SM, Sanders KM, Fujino MA. Novel human and mouse genes encoding a shank-interacting protein and its upregulation in gastric fundus of W/WV mouse. J. Gastroenterol. Hepatol 2003;18:712–8.

15. Riazi MA, Brinkman-Mills P, Johnson A, et al. Identification of a putative regulatory subunit of a calcium-activated potassium channel in the dup(3q) syndrome region and a related sequence on 22q11.2. Genomics 1999;62:90–4.

16. Chiang PW, Wang S, Smithivas P, et al. Identification and analysis of the human and murine putative chromatin structure regulator SUPT6H and Supt6h. Genomics 1996;34:328–33.

17. Nepomuceno RR, Henschen-Edman AH, Burgess WH, Tenner AJ. cDNA cloning and primary structure analysis of C1qR(P), the human C1q/MBL/SPA receptor that mediates enhanced phagocytosis in vitro. Immunity 1997;6:119–29.

18. Totoki Y, Toyoda A, Takeda T, et al. Homo sapiens protein coding cDNA. Submitted (MAR-2005) to the EMBL/GenBank/DDBJ databases.

19. Galiegue S, Jbilo O, Combes T, et al. Cloning and characterization of PRAX-1. A new protein that specifically interacts with the peripheral benzodiazepine receptor. J Biol Chem 1999;274:2938–52.

20. Lahn BT, Tang ZL, Zhou J, et al. Previously uncharacterized histone acetyltransferases implicated in mammalian spermatogenesis. Proc Natl Acad Sci USA 2002;99:8707–12.

21. Clark HF, Gurney AL, Abaya E, et al. The secreted protein discovery initiative (SPDI), a large-scale effort to identify novel human secreted and transmembrane proteins: a bioinformatics assessment. Genome Res 2003;13:2265–70.

22. Zhen YY, Libotte T, Munck M, Noegel AA, Korenbaum E. NUANCE, a giant protein connecting the nucleus and actin cytoskeleton. J Cell Sci 2002;115:3207–22.

23. Johnston H, Kneer J, Chackalaparampil I, Yaciuk P, Chrivia J. Identification of a novel SNF2/SWI2 protein family member, SRCAP, which interacts with CREB-binding protein. J Biol Chem 1999;274:16370–6.

24. Koenig M, Monaco AP, Kunkel LM. The complete sequence of dystrophin predicts a rod-shaped cytoskeletal protein. Cell 1988;53:219–28.

25. Kikuno R, Nagase T, Ishikawa K, et al. Prediction of the coding sequences of unidentified human genes. XIV. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro. DNA Res 1999;6:197–205.

26. Crackower MA, Sinasac DS, Xia J, et al. Cloning and characterization of two cytoplasmic dynein intermediate chain genes in mouse and human.Genomics 1999;55:257–67.

27. The MGC Project Team The status, quality, and expansion of the NIH full-length cDNA project: the Mammalian Gene Collection (MGC). Genome Res 2004;14:2121–7.

28. Meyerhardt JA, Look AT, Bigner SH, Fearon ER. Identification and characterization of neogenin, a DCC-related gene. Oncogene 1997;14:1129–36.

29. Boucrot E, Henry T, Borg JP, Gorvel JP, Meresse S. The intracellular fate of Salmonella depends on the recruitment of kinesin. Science 2005;308:1174–8.

30. Kertesz N, Samson J, Debacker C, Wu H, Labastie MC. Cloning and characterization of human and mouse SNRK sucrose non-fermenting protein (SNF-1)-related kinases. Gene 2002;294:13–24.

31. Sasada R, Ono Y, Taniyama Y, Shing Y, Folkman J, Igarashi K. Cloning and expression of cDNA encoding human betacellulin, a new member of the EGF family. Biochem Biophys Res Commun 1993;190:1173–9.

32. Matuoka K, Miki H, Takahashi K, Takenawa T. A novel ligand for an SH3 domain of the adaptor protein Nck bears an SH2 domain and nuclear signaling motifs. Biochem Biophys Res Commun 1997;239:488–92.

33. Jilani A, Ramotar D, Slack C, et al. Molecular cloning of the human gene, PNKP, encoding a polynucleotide kinase 3'-phosphatase and evidence for its role in repair of DNA strand breaks caused by oxidative damage. J Biol Chem 1999;274:24176–86.

34 Matsuoka S, Ballif BA, Smogorzewska A, et al. ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. Science 2007;316:1160–6.

35. Fujita F, Taniguchi Y, Kato T, et al. Identification of NAP1, a regulatory subunit of IkappaB kinase-related kinases that potentiates NF-kappaB signaling. Mol Cell Biol 2003;23:7780–93.

36. Neufeld EJ, Skalnik DG, Lievens PMJ, Orkin SH. Human CCAAT displacement protein is homologous to the Drosophila homeoprotein, cut. Nat Genet 1992;1:50–5.

37. Raabe T, Manley JL. A human homologue of the Escherichia coli DnaJ heat-shock protein. Nucleic Acids Res. 1991;19:6645–5.

38. Acampora D, D'Esposito M, Faiella A, et al. The human HOX gene family. Nucleic Acids Res 1989;17:10385–402.

39. Langbein L, Heid HW, Moll I, Franke WW. Molecular characterization of the body site-specific human epidermal cytokeratin 9: cDNA cloning, amino acid sequence, and tissue specificity of gene expression. Differentiation 1993;55:57–72.

40. Philipson LH, Hice RE, Schaefer K, et al. Sequence and functional expression in Xenopus oocytes of a human insulinoma and islet potassium channel. Proc Natl Acad Sci USA 1991;88:53–7.

41. Valiente M, Andres-Pons A, Gomar B, et al. Binding of PTEN to specific PDZ domains contributes to PTEN protein stability and phosphorylation by microtubule-associated serine/threonine kinases. J Biol Chem 2005;280:28936–43.

42. Tominaga K, Johmura Y, Nishizuka M, Imagawa M. Fad24, a mammalian homolog of Noc3p, is a positive regulator in adipocyte differentiation. J Cell Sci 2004;117:6217–26.

43. Huang Y, Zhou XM, Zhang PP, et al. Novel human cDNA clones with function of inhibiting cancer cell growth. Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.

44. Matsuda A, Suzuki Y, Honda G, et al. Large-scale identification and characterization of human genes that activate NF-kappaB and MAPK signaling pathways. Oncogene 2003;22:3307–18.

45. Nagase T, Nakayama M, Nakajima D, Kikuno R, Ohara O. Prediction of the coding sequences of unidentified human genes. XX. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro. DNA Res 2001;8:85–95.

46. Nagase T, Kikuno R, Ohara O. Prediction of the coding sequences of unidentified human genes. XXI. The complete sequences of 60 new cDNA clones from brain which code for large proteins. DNA Res 2001;8:179–87.

47. Lefebvre S, Buerglen L, Reboullet S, et al. Identification and characterization of a spinal muscular atrophy-determining gene. Cell 1995;80:155–65.

48. Edgar AJ, Milona MA. Expression of human specificity protein 8 (SP8) in osteoblasts. Submitted (OCT-2002) to the EMBL/GenBank/DDBJ databases.

49. Debant A, Serra-Pages C, Seipel K, et al. The multidomain protein Trio binds the LAR transmembrane tyrosine phosphatase, contains a protein kinase domain, and has separate rac-specific and rho-specifi-57c guanine nucleotide exchange factor domains. Proc Natl Acad Sci USA 1996;93:5466–71.

50. Moniaux N, Junker WM, Singh AP, Jones AM, Batra SK. Characterization of human mucin MUC17. Complete coding sequence and organization. J Biol Chem 2006;281:23676–85.

51. Nguyen TB, Manova K, Capodieci P, et al. Characterization and expression of mammalian cyclin b3, a prepachytene meiotic cyclin. J Biol Chem 2002;277:41960–9.

52. Stenmark H, Vitale G, Ulrich O, Zerial M. Rabaptin-5 is a direct effector of the small GTPase Rab5 in endocytic membrane fusion. Cell 1995;83:423–32.

53. Ross MT, Grafham DV, Coffey AJ, et al. The DNA sequence of the human X chromosome. Nature. 2005;434:325–37.

54. Rush J, Moritz A, Lee KA, et al. Immunoaffinity profiling of tyrosine phosphorylation in cancer cells. Nat Biotechnol 2005;23:94–101.

55. Wunderle VM, Critcher R, Ashworth A, Goodfellow PN. Cloning and characterization of SOX5, a new member of the human SOX gene family. Genomics 1996;36:354–8.

56. Ramesh N, Anton IM, Hartwig JH, Geha RS. WIP, a protein associated with Wiskott-Aldrich syndrome protein, induces actin polymerization and redistribution in lymphoid cells. Proc Natl Acad Sci USA 1997;94:14671–6.

57. Prades C, Arnould I, Annilo T, et al. The human ATP binding cassette gene ABCA13, located on chromosome 7p12.3, encodes a 5058 amino acid protein with an extracellular domain encoded in part by a 4.8-kb conserved exon. Cytogenet Genome Res 2002;98:160–8.

58. Noor A, Windpassinger C, Patel M, et al. CC2D2A, encoding a coiled-coil and C2 domain protein, causes autosomal-recessive mental retardation with retinitis pigmentosa. Am J Hum Genet 2008;82:1011–1018.

59. Sakai N, Inui K, Fujii N, et al. Krabbe disease: isolation and characterization of a full-length cDNA for human galactocerebrosidase. Biochem Biophys Res Commun 1994;198:485–91.

60. Galaktionov KI, Beach D. Specific activation of cdc25 tyrosine phosphatases by B-type cyclins: evidence for multiple roles of mitotic cyclins. Cell 1991;67:1181–94.

61. Walter DM, Paul KS, Waters MG. Purification and characterization of a novel 13 S hetero-oligomeric protein complex that stimulates in vitro Golgi transport. J Biol Chem 1998;273:29565–76.

62. Hästbacka J, de la Chapelle A, Mahtani MM, et al. The diastrophic dysplasia gene encodes a novel sulfate transporter: positional cloning by fine-structure linkage disequilibrium mapping. Cell 1994;78:1073–87.

63. O'Neill RE, Palese P. NPI-1, the human homolog of SRP-1, interacts with influenza virus nucleoprotein. Virology 1995;206:116–25.

64. Köhler M, Ansieau S, Prehn S, Leutz A, Haller H, Hartmann E. Cloning of two novel human importin-alpha subunits and analysis of the expression pattern of the importin-alpha protein family. FEBS Lett 1997;417:104–8.

65. Köhler M, Speck C, Christiansen M, et al. Evidence for distinct substrate specificities of importin alpha family members in nuclear protein import. Mol Cell Biol 1999;19:7782–91.

66. Melis D, Havelaar AC, Verbeek E, et al. NPT4, a new microsomal phosphate transporter: mutation analysis in glycogen storage disease type Ic. J Inherit Metab Dis 2004;27:725–33.

67. Buyse IM, Shao G, Huang S. The retinoblastoma protein binds to RIZ, a zinc-finger protein that shares an epitope with the adenovirus E1A protein. Proc Natl Acad Sci USA 1995;92:4467–71.

68. Fattaey AR, Helin K, Dembski MS, et al. Characterization of the retinoblastoma binding proteins RBP1 and RBP2. Oncogene 1993;8:3149–56.

69. Shiomi T, Harada YN, Saito T, Shiomi N, Okuno Y, Yamaizumi M. An ERCC5 gene with homology to yeast RAD2 is involved in group G Xeroderma pigmentosum. Mutat Res 1994;314:167–75.

70. Heisterkamp N, Groffen J, Warburton D, Sneddon TP. The human gamma-glutamyltransferase gene family. Hum Genet 2008;123:321–32.

71. Mural RJ, Istrail S, Sutton G, et al. Submitted (SEP-2005) to the EMBL/GenBank/DDBJ databases.