APPENDIX A

Vega Gene ID	HUGO Gene ID	Gene Description	Summary
OTTHUMG00000033045	ACOT8	acyl-CoA thioesterase 8	The protein encoded by this gene is a peroxisomal thioesterase that appears to be involved more in the oxidation of fatty acids rather than in their formation. The encoded protein can bind to the human immunodeficiency virus-1 protein Nef, and mediate Nef-induced down-regulation of CD4 in T-cells. Multiple transcript variants encoding several different isoforms have been found for this gene.
OTTHUMG00000033081	ADA	adenosine deaminase	Adenosine deaminase catalyzes the hydrolysis of adenosine to inosine. ADA deficiency causes one form of severe combined immunodeficiency disease (SCID), in which there is dysfunction of both B and T lymphocytes with impaired cellular immunity and decreased production of immunoglobulins.
OTTHUMG00000032687	ARFGEF2	ADP-ribosylation factor guanine nucleotide-exchange factor 2 (brefeldin A-inhibited)	ADP-ribosylation factors (ARFs) play an important role in intracellular vesicular trafficking. The protein encoded by this gene is involved in the activation of ARFs by accelerating replacement of bound GDP with GTP and is involved in Golgi transport. It contains a Sec7 domain, which may be responsible for its guanine-nucleotide exchange activity and also brefeldin A inhibition.
OTTHUMG00000033086	B4GALT5	UDP-Gal:betaGlcNAc beta 1,4-galactosyltransferase, polypeptide 5	This gene is one of seven beta-1,4-galactosyltransferase (beta4GalT) genes. They encode type II membrane-bound glycoproteins that appear to have exclusive specificity for the donor substrate UDP-galactose; all transfer galactose in a beta1,4 linkage to similar acceptor sugars: GlcNAc, Glc, and Xyl. Each beta4GalT has a distinct function in the biosynthesis of different glycoconjugates and saccharide structures. As type II membrane proteins, they have an N-terminal hydrophobic signal sequence that directs the protein to the Golgi apparatus and which then remains uncleaved to function as a transmembrane anchor. By sequence similarity, the beta4GalTs form four groups: beta4GalT1 and beta4GalT2, beta4GalT3 and beta4GalT4, beta4GalT5 and beta4GalT6, and beta4GalT7. The function of the enzyme encoded by this gene is not clear. This gene was previously designated as B4GALT4 but was renamed to B4GALT5. In the literature it is also referred to as beta4GalT2.
OTTHUMG00000032582	C20orf10	chromosome 20 open reading frame 10	
OTTHUMG00000032517	C20orf100	chromosome 20 open reading frame 100	
OTTHUMG00000032518	C20orf111	chromosome 20 open reading frame 111	
OTTHUMG00000032553	C20orf119	chromsome 20 open reading frame 119	
OTTHUMG00000032536	C20orf121	chromosome 20 open reading frame 121	
OTTHUMG00000032652	C20orf123	chromosome 20 open reading frame 123	
OTTHUMG00000032522	C20orf142	chromosome 20 open reading frame 142	
OTTHUMG00000032647	C20orf157	chromosome 20 open reading frame 157	

OTTHUMG00000032624	SNX21 (C20ORF161)	sorting nexin family member 21	This gene encodes a member of the sorting nexin family. Members of this family contain a phox (PX) domain, which is a phosphoinositide binding domain, and are involved in intracellular trafficking. This protein does not contain a coiled coil region, like some family members. The specific function of this protein has not been determined. Multiple transcript variants encoding distinct isoforms have been identified for this gene.
OTTHUMG00000032628	C20orf165	chromosome 20 open reading frame 165	
OTTHUMG00000032611	C20orf168	chromosome 20 open reading frame 168	
OTTHUMG00000032576	DBNDD2 (C20ORF35)	dysbindin (dystrobrevin binding protein 1) domain containing 2	
OTTHUMG00000032533	C20orf62	chromosome 20 open reading frame 62	
OTTHUMG00000032510	FAM112A (RP5-10496-16.1)	family with sequence similarity 112, member A	
OTTHUMG00000032635	C20orf67	Chromosome 20 open reading frame 67	
OTTHUMG00000033053	CD40	CD40 molecule, TNF receptor superfamily member 5	The protein encoded by this gene is a member of the TNF-receptor superfamily. This receptor has been found to be essential in mediating a broad variety of immune and inflammatory responses including T cell-dependent immunoglobulin class switching, memory B cell development, and germinal center formation. AT-hook transcription factor AKNA is reported to coordinately regulate the expression of this receptor and its ligand, which may be important for homotypic cell interactions. Adaptor protein TNFR2 interacts with this receptor and serves as a mediator of the signal transduction. The interaction of this receptor and its ligand is found to be necessary for amyloid-beta-induced microglial activation, and thus is thought to be an early event in Alzheimer disease pathogenesis. Two alternatively spliced transcript variants of this gene encoding distinct isoforms have been reported.
OTTHUMG00000033073	CDH22	cadherin-like 22	This gene is a member of the cadherin superfamily. The gene product is composed of five cadherin repeat domains and a cytoplasmic tail similar to the highly conserved cytoplasmic region of classical cadherins. Expressed predominantly in the brain, this putative calcium-dependent cell adhesion protein may play an important role in morphogenesis and tissue formation in neural and non-neural cells during development and maintenance of the brain and neuroendocrine organs.
OTTHUMG00000032487	CHD6	chromodomain helicase DNA binding protein 6	Chromosomal DNA of eukaryotic cells is compacted by nuclear proteins to form chromatin, an organized nucleoprotein structure that can inhibit gene expression. Several multisubunit protein complexes exist to remodel the chromatin to allow patterns of cell type-specific gene expression. The protein encoded by this gene is thought to be a core member of one or more of these complexes. The encoded protein, which is a member of the SNF2/RAD54 helicase family, contains two chromodomains, a helicase domain, and an ATPase domain.

OTTHUMG00000033046	CSE1L	CSE1 chromosome segregation 1-like (yeast)	Proteins that carry a nuclear localization signal (NLS) are transported into the nucleus by the importin-alpha/beta heterodimer. Importin-alpha binds the NLS, while importin-beta mediates translocation through the nuclear pore complex. After translocation, RanGTP binds importin-beta and displaces importin-alpha. Importin-alpha must then be returned to the cytoplasm, leaving the NLS protein behind. The protein encoded by this gene binds strongly to NLS-free importin-alpha, and this binding is released in the cytoplasm by the combined action of RANBP1 and RANGAP1. In addition, the encoded protein may play a role both in apoptosis and in cell proliferation.
OTTHUMG00000033072	DDX27	DEAD (Asp-Glu-Ala-Asp) box polypeptide 27	DEAD box proteins, characterized by the conserved motif Asp-Glu-Ala-Asp (DEAD), are putative RNA helicases. They are implicated in a number of cellular processes involving alteration of RNA secary structure such as translation initiation, nuclear and mitochondrial splicing, and ribosome and spliceosome assembly. Based on their distribution patterns, some members of this family are believed to be involved in embryogenesis, spermatogenesis, and cellular growth and division. This gene encodes a DEAD box protein, the function of which has not been determined.
OTTHUMG00000032610	DNTTIP1	deoxynucleotidyltransferase, terminal, interacting protein 1	
OTTHUMG00000033070	ELMO2	engulfment and cell motility 2	The protein encoded by this gene interacts with the dedicator of cyto-kinesis 1 protein. Similarity to a C. elegans protein suggests that this protein may function in phagocytosis of apoptotic cells and in cell migration. Alternative splicing results in multiple transcript variants encoding the same protein.
OTTHUMG00000046304	EMILIN3	elastin microfibril interfacer 3	
OTTHUMG00000033041	EYA2	eyes absent homolog 2 (Drosophila)	This gene encodes a member of the eyes absent (EYA) family of proteins. The encoded protein may be post-translationally modified and may play a role in eye development. A similar protein in mice can act as a transcriptional activator. Five transcript variants encoding three distinct isoforms have been identified for this gene.
OTTHUMG00000032530	GDAP1L1	ganglioside-induced differentiation-associated protein 1-like 1	The ganglioside GD3 synthase causes cell differentiation with neurite sprouting when transfected into the mouse neuroblastoma cell line Neuro2a. After differentiation, the expression of several genes is upregulated, including one that encodes a protein termed ganglioside-induced differentiation-associated protein 1 (Gdap1). A similar gene was found in humans, and mutations in the human gene are associated with Charcot-Marie-Tooth type 4A disease. The protein encoded by this gene is similar in sequence to the human GDAP1 protein.
OTTHUMG00000032531	HNF4A	hepatocyte nuclear factor 4, alpha	The protein encoded by this gene is a nuclear transcription factor which binds DNA as a homodimer. The encoded protein controls the expression of several genes, including hepatocyte nuclear factor 1 alpha, a transcription factor which regulates the expression of several hepatic genes. This gene may play a role in development of the liver, kidney, and intestines. Mutations in this gene have been associated with monogenic autosomal dominant non-insulin-dependent diabetes mellitus type I. Alternative splicing of this

			gene results in multiple transcript variants.
OTTHUMG00000032513	IFT52	intraflagellar transport 52 homolog (Chlamydomonas)	
OTTHUMG00000033037	JPH2	junctophilin 2	Junctional complexes between the plasma membrane and endoplasmic/sarcoplasmic reticulum are a common feature of all excitable cell types and mediate cross talk between cell surface and intracellular ion channels. The protein encoded by this gene is a component of junctional complexes and is composed of a C-terminal hydrophobic segment spanning the endoplasmic/sarcoplasmic reticulum membrane and a remaining cytoplasmic domain that shows specific affinity for the plasma membrane. This gene is a member of the junctophilin gene family. Alternative splicing has been observed at this locus and two variants encoding distinct isoforms are described.
OTTHUMG00000033051	KCNB1	potassium voltage-gated channel, Shab-related subfamily, member 1	Voltage-gated potassium (Kv) channels represent the most complex class of voltage-gated ion channels from both functional and structural standpoints. Their diverse functions include regulating neurotransmitter release, heart rate, insulin secretion, neuronal excitability, epithelial electrolyte transport, smooth muscle contraction, and cell volume. Four sequence-related potassium channel genes - shaker, shaw, shab, and shal - have been identified in Drosophila, and each has been shown to have human homolog(s). This gene encodes a member of the potassium channel, voltage-gated, shab-related subfamily. This member is a delayed rectifier potassium channel and its activity is modulated by some other family members.
OTTHUMG00000032544	KCNK15	potassium channel, subfamily K, member 15	This gene encodes one of the members of the superfamily of potassium channel proteins containing two pore-forming P domains. The product of this gene has not been shown to be a functional channel, however, it may require other non-pore-forming proteins for activity.
OTTHUMG00000033079	KCNS1	potassium voltage-gated channel, delayed-rectifier, subfamily S, member 1	Voltage-gated potassium channels form the largest and most diversified class of ion channels and are present in both excitable and nonexcitable cells. Their main functions are associated with the regulation of the resting membrane potential and the control of the shape and frequency of action potentials. The alpha subunits are of 2 types: those that are functional by themselves and those that are electrically silent but capable of modulating the activity of specific functional alpha subunits. The protein encoded by this gene is not functional by itself but can form heteromultimers with member 1 and with member 2 (and possibly other members) of the Shab-related subfamily of potassium voltage-gated channel proteins. This gene belongs to the S subfamily of the potassium channel family.
OTTHUMG00000032696	ZNFX1 (KIAA1404)	zinc finger, NFX1-type containing 1	
OTTHUMG00000032503	L3MBTL	l(3)mbt-like (Drosophila)	This gene encodes the homolog of a protein identified in Drosophila as a suppressor of malignant transformation of neuroblasts and ganglion-mother cells in the optic centers of the brain. This gene product is localized to condensed chromosomes in mitotic cells.

			Overexpression of this gene in a glioma cell line results in improper nuclear segregation and cytokinesis producing multinucleated cells. Two transcripts have been identified for this gene.
OTTHUMG00000033056	LPIN3	lipin 3	Humans lipodystrophy is characterized by loss of body fat, fatty liver, hypertriglyceridemia, and insulin resistance. Mice carrying mutations in the fatty liver dystrophy (fld) gene have similar phenotypes. Through positional cloning, the mouse gene responsible for fatty liver dystrophy was isolated and designated Lpin1. The nuclear protein encoded by Lpin1 was named lipin. Lpin1 mRNA was expressed at high levels in adipose tissue and was induced during differentiation of preadipocytes. These results indicated that lipin is required for normal adipose tissue development and provided a candidate gene for human lipodystrophy. Through database searches, mouse and human EST and genomic sequences with similarities to Lpin1 were identified. These included two related mouse genes (Lpin2 and Lpin3) and three human homologs (LPIN1, LPIN2, and LPIN3). Human LPIN1 gene has been mapped to 2p25.; linkages of fat mass and serum leptin levels to this same region have been noted. Human LPIN2 and LPIN3 mapped to chromosomes 1
OTTHUMG00000033052	MAFB	v-maf musculoaponeurotic fibrosarcoma oncogene homolog B (avian)	The protein encoded by this gene is a basic leucine zipper (bZIP) transcription factor that plays an important role in the regulation of lineage-specific hematopoiesis. The encoded nuclear protein represses ETS1-mediated transcription of erythroid-specific genes in myeloid cells. This gene contains no introns.
OTTHUMG00000033043	MATN4	matrilin 4	This gene encodes a member of von Willebrand factor A domain containing protein family. The proteins of this family are thought to be involved in the formation of filamentous networks in the extracellular matrices of various tissues. The specific function of this gene product has not yet been determined. Three alternatively spliced variants have been described.
OTTHUMG00000033044	MMP9	matrix metalloproteinase 9 (gelatinase B, 92kDa gelatinase, 92kDa type IV collagenase)	Proteins of the matrix metalloproteinase (MMP) family are involved in the breakdown of extracellular matrix in normal physiological processes, such as embryonic development, reproduction, and tissue remodeling, as well as in disease processes, such as arthritis and metastasis. Most MMP's are secreted as inactive proproteins which are activated when cleaved by extracellular proteinases. The enzyme encoded by this gene degrades type IV and V collagens. Studies in rhesus monkeys suggest that the enzyme is involved in IL-8-induced mobilization of hematopoietic progenitor cells from bone marrow, and murine studies suggest a role in tumor-associated tissue remodeling.
OTTHUMG00000033062	MYBL2	v-myb myeloblastosis viral oncogene homolog (avian)-like 2	The protein encoded by this gene, a member of the MYB family of transcription factor genes, is a nuclear protein involved in cell cycle progression. The encoded protein is phosphorylated by cyclin A/cyclin-dependent kinase 2 during the S-phase of the cell cycle and possesses both activator and repressor activities. It has been shown to activate the cell division cycle 2, cyclin D1, and insulin-like growth factor-binding protein 5 genes. Transcript variants may exist for this gene, but their full-length natures have not

			been determined.
OTTHUMG00000033061	NCOA3	nuclear receptor coactivator 3	The protein encoded by this gene is a nuclear receptor coactivator that interacts with nuclear hormone receptors to enhance their transcriptional activator functions. The encoded protein has histone acetyltransferase activity and recruits p300/CBP-associated factor and CREB binding protein as part of a multisubunit coactivation complex. This protein is initially found in the cytoplasm but is translocated into the nucleus upon phosphorylation. Two transcript variants encoding different isoforms have been found for this gene. In addition, a polymorphic repeat region is found in the C-terminus of the encoded protein.
OTTHUMG00000032639	NCOA5	nuclear receptor coactivator 5	This gene encodes a coregulator for the alpha and beta estrogen receptors and the orphan nuclear receptor NR1D2. The protein localizes to the nucleus, and is thought to have both coactivator and corepressor functions. Its interaction with nuclear receptors is independent of the AF2 domain on the receptors, which is known to regulate interaction with other coreceptors. Two alternatively spliced transcript variants for this gene have been described. However, the full length nature of one of the variants has not been determined.
OTTHUMG00000032626	NEURL2	neuralized-like 2 (Drosophila)	
OTTHUMG00000032567	PI3	protease inhibitor 3, skin-derived (SKALP)	This gene encodes an elastase-specific inhibitor, which contains a WAP-type four-disulfide core (WFDC) domain, and is thus a member of the WFDC domain family. Most WFDC gene members are localized to chromosome 20q12-q13 in two clusters: centromeric and telomeric. This gene belongs to the centromeric cluster.
OTTHUMG00000032574	PIGT	phosphatidylinositol glycan anchor biosynthesis, class T	This gene encodes a protein that is involved in glycosylphosphatidylinositol (GPI)-anchor biosynthesis. The GPI-anchor is a glycolipid found on many blood cells and serves to anchor proteins to the cell surface. This protein is an essential component of the multisubunit enzyme, GPI transamidase. GPI transamidase mediates GPI anchoring in the endoplasmic reticulum, by catalyzing the transfer of fully assembled GPI units to proteins.
OTTHUMG00000033065	PKIG	protein kinase (cAMP-dependent, catalytic) inhibitor gamma	The protein encoded by this gene is a member of the cAMP-dependent protein kinase (PKA) inhibitor family. Studies of a similar protein in mice suggest that this protein acts as a potent competitive PKA inhibitor, and is a predominant form of PKA inhibitors in various tissues. Three alternatively spliced transcript variants encoding the same protein have been reported.
OTTHUMG00000033082	PLCG1	phospholipase C, gamma 1	The protein encoded by this gene catalyzes the formation of inositol 1,4,5-trisphosphate and diacylglycerol from phosphatidylinositol 4,5-bisphosphate. This reaction uses calcium as a cofactor and plays an important role in the intracellular transduction of receptor-mediated tyrosine kinase activators. For example, when activated by SRC, the encoded protein causes the Ras guanine nucleotide exchange factor RasGRP1 to translocate to the Golgi, where it activates Ras. Also, this protein has been shown to be a major substrate for heparin-binding growth factor 1 (acidic fibroblast growth factor)-

			activated tyrosine kinase. Two transcript variants encoding different isoforms have been found for this gene.
OTTHUMG00000033047	PLTP	phospholipid transfer protein	The protein encoded by this gene is one of at least two lipid transfer proteins found in human plasma. The encoded protein transfers phospholipids from triglyceride-rich lipoproteins to high density lipoprotein (HDL). In addition to regulating the size of HDL particles, this protein may be involved in cholesterol metabolism. At least two transcript variants encoding different isoforms have been found for this gene.
OTTHUMG00000033078	PPGB	protective protein for beta-galactosidase (galactosialidosis)	This gene encodes a glycoprotein which associates with lysosomal enzymes beta-galactosidase and neuraminidase to form a complex of high molecular weight multimers. The formation of this complex provides a protective role for stability and activity. Deficiencies in this gene are linked to multiple forms of galactosialidosis.
OTTHUMG00000032667	PRKCBP1	protein kinase C binding protein 1	The protein encoded by this gene is a receptor for activated C-kinase (RACK) protein. The encoded protein has been shown to bind in vitro to activated protein kinase C beta I. In addition, this protein is a cutaneous T-cell lymphoma-associated antigen. Finally, the protein contains a bromodomain and two zinc fingers, and is thought to be a transcriptional regulator. Multiple transcript variants encoding several different isoforms have been found for this gene.
OTTHUMG00000033077	PTGIS	prostaglandin I2 (prostacyclin) synthase	This gene encodes a member of the cytochrome P450 superfamily of enzymes. The cytochrome P450 proteins are monooxygenases which catalyze many reactions involved in drug metabolism and synthesis of cholesterol, steroids and other lipids. However, this protein is considered a member of the cytochrome P450 superfamily on the basis of sequence similarity rather than functional similarity. This endoplasmic reticulum membrane protein catalyzes the conversion of prostglandin H2 to prostacyclin (prostaglandin I2), a potent vasodilator and inhibitor of platelet aggregation. An imbalance of prostacyclin and its physiological antagonist thromboxane A2 contribute to the development of myocardial infarction, stroke, and atherosclerosis.
OTTHUMG00000033040	PTPRT	protein tyrosine phosphatase, receptor type, T	The protein encoded by this gene is a member of the protein tyrosine phosphatase (PTP) family. PTPs are known to be signaling molecules that regulate a variety of cellular processes including cell growth, differentiation, mitotic cycle, and oncogenic transformation. This PTP possesses an extracellular region, a single transmembrane region, and two tandem intracellular catalytic domains, and thus represents a receptor-type PTP. The extracellular region contains a meprin-A5 antigen-PTP (MAM) domain, Ig-like and fibronectin type III-like repeats. The protein domain structure and the expression pattern of the mouse counterpart of this PTP suggest its roles in both signal transduction and cellular adhesion in the central nervous system. Two alternatively spliced transcript variants of this gene, which encode distinct proteins, have been reported.
OTTHUMG00000032524	R3HDML	R3H domain (binds single-stranded nucleic acids) containing-like	

OTTHUMG00000033055	RBPSUHL	recombining binding protein suppressor of hairless (Drosophila)-like	In mouse, recombining binding protein L (RBP-L) is a transcription factor that binds to DNA sequences almost identical to that bound by the Notch receptor signalling pathway transcription factor RBP-J. However, unlike RBP-J, RBP-L does not interact with Notch receptors. RBP-L has been shown to activate transcription in concert with Epstein-Barr virus nuclear antigen-2 (EBNA2). The protein encoded by this gene is similar in sequence to the mouse RPB-L protein and Drosophila suppressor of hairless protein.
OTTHUMG00000032546	RIMS4	regulating synaptic membrane exocytosis 4	
OTTHUMG00000032685	PREX1 (RP11-269H4.1)	phosphatidylinositol 3,4,5-trisphosphate-dependent RAC exchanger 1	The protein encoded by this gene acts as a guanine nucleotide exchange factor for the RHO family of small GTP-binding proteins (RACs). It has been shown to bind to and activate RAC1 by exchanging bound GDP for free GTP. The encoded protein, which is found mainly in the cytoplasm, is activated by phosphatidylinositol-3,4,5-trisphosphate and the beta-gamma subunits of heterotrimeric G proteins.
OTTHUMG00000032675	SULF2	sulfatase 2	Heparan sulfate proteoglycans (HSPGs) act as coreceptors for numerous heparin-binding growth factors and cytokines and are involved in cell signaling. Heparan sulfate 6-O-endosulfatases, such as SULF2, selectively remove 6-O-sulfate groups from heparan sulfate. This activity modulates the effects of heparan sulfate by altering binding sites for signaling molecules (Dai et al., 2005).[supplied by OMIM]
OTTHUMG00000033083	SDC4	syndecan 4 (amphiglycan, ryudocan)	The protein encoded by this gene is a transmembrane (type I) heparan sulfate proteoglycan that functions as a receptor in intracellular signaling. The encoded protein is found as a homodimer and is a member of the syndecan proteoglycan family. This gene is found on chromosome 20, while a pseudogene has been found on chromosome 22.
OTTHUMG00000032565	SEMG1	semenogelin I	The protein encoded by this gene is the predominant protein in semen. The encoded secreted protein is involved in the formation of a gel matrix that encases ejaculated spermatozoa. The prostate-specific antigen (PSA) protease processes this protein into smaller peptides, with each possibly having a separate function. The proteolysis process breaks down the gel matrix and allows the spermatozoa to move more freely. Two transcript variants encoding different isoforms have been found for this gene.
OTTHUMG00000032566	SEMG2	semenogelin II	The secreted protein encoded by this gene is involved in the formation of a gel matrix that encases ejaculated spermatozoa. Proteolysis by the prostate-specific antigen (PSA) breaks down the gel matrix and allows the spermatozoa to move more freely. The encoded protein is found in lesser abundance than a similar semenogelin protein. The genes encoding these two semenogelin proteins are found in a cluster on chromosome 20.
OTTHUMG00000032502	SFRS6	splicing factor, arginine/serine-rich 6	The protein encoded by this gene is involved in mRNA splicing and may play a role in the determination of alternative splicing. The encoded nuclear protein belongs to the splicing factor SR family and has been shown to bind with and modulate another

			member of the family, SFRS12.
OTTHUMG00000033054	SGK2	serum/glucocorticoid regulated kinase 2	This gene encodes a serine/threonine protein kinase. Although this gene product is similar to serum- and glucocorticoid-induced protein kinase (SGK), this gene is not induced by serum or glucocorticoids. This gene is induced in response to signals that activate phosphatidylinositol 3-kinase, which is also true for SGK. Two alternate transcripts encoding two different isoforms have been described.
OTTHUMG00000032638	SLC12A5	solute carrier family 12, (potassium-chloride transporter) member 5	K-Cl cotransporters are proteins that lower intracellular chloride concentrations below the electrochemical equilibrium potential. The protein encoded by this gene is an integral membrane K-Cl cotransporter that can function in either a net efflux or influx pathway, depending on the chemical concentration gradients of potassium and chloride. The encoded protein can act as a homomultimer, or as a heteromultimer with other K-Cl cotransporters, to maintain chloride homeostasis in neurons.
OTTHUMG00000033042	SLC13A3	solute carrier family 13 (sodium-dependent dicarboxylate transporter), member 3	Mammalian sodium-dicarboxylate cotransporters transport succinate and other Krebs cycle intermediates. They fall into 2 categories based on their substrate affinity: low affinity and high affinity. Both the low- and high-affinity transporters play an important role in the handling of citrate by the kidneys. The protein encoded by this gene represents the high-affinity form. Alternatively spliced transcript variants encoding different isoforms have been found for this gene, although the full-length nature of some of them have not been characterized yet.
OTTHUMG00000032657	SLC2A10	solute carrier family 2 (facilitated glucose transporter), member 10	SLC2A10 is a member of the facilitative glucose transporter family, which plays a significant role in maintaining glucose homeostasis.[supplied by OMIM]
OTTHUMG00000033050	SLC35C2	solute carrier family 35, member C2	Oxygenation levels play an important role in the regulation of cellular invasiveness which occurs during early implantation when the trophoblast cells invade the uterus as well as during tumour progression and metastasis. This gene, which is regulated by oxygen tension, is induced in hypoxic trophoblast cells and is overexpressed in ovarian cancer. Two protein isoforms are encoded by transcript variants of this gene.
OTTHUMG00000032710	SLC9A8	solute carrier family 9 (sodium\/hydrogen exchanger), isoform 8	
OTTHUMG00000033075	SLPI	secretory leukocyte peptidase inhibitor	This gene encodes a secreted inhibitor which protects epithelial tissues from serine proteases. It is found in various secretions including seminal plasma, cervical mucus, and bronchial secretions, and has affinity for trypsin, leukocyte elastase, and cathepsin G. Its inhibitory effect contributes to the immune response by protecting epithelial surfaces from attack by endogenous proteolytic enzymes; the protein is also thought to have broad-spectrum anti-biotic activity.
OTTHUMG00000033048	SNAI1	snail homolog 1 (Drosophila)	The Drosophila embryonic protein snail is a zinc finger transcriptional repressor which downregulates the expression of ectodermal genes within the mesoderm. The nuclear protein encoded by this gene is structurally similar to the Drosophila snail protein, and is also thought to be critical for mesoderm formation in the developing embryo. At least two variants of a similar processed pseudogene have been found on chromosome 2.

OTTHUMG00000032704	SPATA2	spermatogenesis associated 2	
OTTHUMG00000032588	SPINLW1	serine protease inhibitor-like, with Kunitz and WAP domains 1 (eppin)	This gene encodes an epididymal protease inhibitor, which contains both kunitz-type and WAP-type four-disulfide core (WFDC) protease inhibitor consensus sequences. Most WFDC genes are localized to chromosome 20q12-q13 in two clusters: centromeric and telomeric. This gene is a member of the WFDC gene family and belongs to the telomeric cluster. Alternatively spliced transcript variants encoding distinct isoforms have been found for this gene.
OTTHUMG00000032585	SPINT3	serine protease inhibitor, Kunitz type, 3	
OTTHUMG00000032604	SPINT4	chromosome 20 open reading frame 137	
OTTHUMG00000032691	STAU1	staufen, RNA binding protein, homolog 1 (Drosophila)	Staufen is a member of the family of double-stranded RNA (dsRNA)-binding proteins involved in the transport and/or localization of mRNAs to different subcellular compartments and/or organelles. These proteins are characterized by the presence of multiple dsRNA-binding domains which are required to bind RNAs having double-stranded secary structures. The human homologue of staufen encoded by STAU, in addition contains a microtubule- binding domain similar to that of microtubule-associated protein 1B, and binds tubulin. The STAU gene product has been shown to be present in the cytoplasm in association with the rough endoplasmic reticulum (RER), implicating this protein in the transport of mRNA via the microtubule network to the RER, the site of translation. Five transcript variants resulting from alternative splicing of STAU gene and encoding three isoforms have been described. Three of these variants encode the same isoform, however, differ in their 5'UTR.
OTTHUMG00000033059	STK4	serine/threonine kinase 4	The protein encoded by this gene is a cytoplasmic kinase that is structurally similar to the yeast Ste20p kinase, which acts upstream of the stress-induced mitogen-activated protein kinase cascade. The encoded protein can phosphorylate myelin basic protein and undergoes autophosphorylation. A caspase-cleaved fragment of the encoded protein has been shown to be capable of phosphorylating histone H2B. The particular phosphorylation catalyzed by this protein has been correlated with apoptosis, and it's possible that this protein induces the chromatin condensation observed in this process.
OTTHUMG00000033087	SERINC3 (TDE1)	serine incorporator 3	
OTTHUMG00000032623	TNNC2	troponin C type 2 (fast)	Troponin (Tn), a key protein complex in the regulation of striated muscle contraction, is composed of 3 subunits. The Tn-I subunit inhibits actomyosin ATPase, the Tn-T subunit binds tropomyosin and Tn-C, while the Tn-C subunit binds calcium and overcomes the inhibitory action of the troponin complex on actin filaments. The protein encoded by this gene is the Tn-C subunit.
OTTHUMG00000032552	TOMM34	translocase of outer mitochondrial membrane 34	The protein encoded by this gene is involved in the import of precursor proteins into mitochondria. The encoded protein has a chaperone-like activity, binding the mature portion of unfolded proteins and aiding their import into mitochondria. This protein, which is found in the cytoplasm and sometimes associated with the outer mitochondrial

			membrane, has a weak ATPase activity and contains 6 TPR repeats.
OTTHUMG00000033057	TOP1	topoisomerase (DNA) I	This gene encodes a DNA topoisomerase, an enzyme that controls and alters the topologic states of DNA during transcription. This enzyme catalyzes the transient breaking and rejoining of a single strand of DNA which allows the strands to pass through one another, thus altering the topology of DNA. This gene is localized to chromosome 20 and has pseudogenes which reside on chromosomes 1 and 22.
OTTHUMG00000085887	TP53RK	TP53 regulating kinase	
OTTHUMG00000033038	UBE2C	ubiquitin-conjugating enzyme E2C	The modification of proteins with ubiquitin is an important cellular mechanism for targeting abnormal or short-lived proteins for degradation. Ubiquitination involves at least three classes of enzymes: ubiquitin-activating enzymes, or E1s, ubiquitin-conjugating enzymes, or E2s, and ubiquitin-protein ligases, or E3s. This gene encodes a member of the E2 ubiquitin-conjugating enzyme family. This enzyme is required for the destruction of mitotic cyclins and for cell cycle progression. Multiple alternatively spliced transcript variants have been found for this gene, but the full-length nature of some variants has not been defined.
OTTHUMG00000033085	UBE2V1	ubiquitin-conjugating enzyme E2 variant 1	Ubiquitin-conjugating E2 enzyme variant proteins constitute a distinct subfamily within the E2 protein family. They have sequence similarity to other ubiquitin-conjugating enzymes but lack the conserved cysteine residue that is critical for the catalytic activity of E2s. The protein encoded by this gene is located in the nucleus and can cause transcriptional activation of the human FOS proto-oncogene. It is thought to be involved in the control of differentiation by altering cell cycle behavior. Multiple alternatively spliced transcripts encoding different isoforms have been described for this gene. A pseudogene has been identified which is also located on chromosome 20. Co-transcription of this gene and the neighboring upstream gene generates a rare transcript (Kua-UEV), which encodes a fusion protein comprised of sequence sharing identity with each individual gene product.
OTTHUMG00000046331	WFDC10A	WAP four-disulfide core domain 10A	This gene encodes a member of the WAP-type four-disulfide core (WFDC) domain family. The WFDC domain, or WAP signature motif, contains eight cysteines forming four disulfide bonds at the core of the protein, and functions as a protease inhibitor. Most WFDC gene members are localized to chromosome 20q12-q13 in two clusters: centromeric and telomeric. This gene belongs to the telomeric cluster.
OTTHUMG00000046334	WFDC10B	WAP four-disulfide core domain 10B	This gene encodes a member of the WAP-type four-disulfide core (WFDC) domain family. The WFDC domain, or WAP signature motif, contains eight cysteines forming four disulfide bonds at the core of the protein, and functions as a protease inhibitor. Most WFDC gene members are localized to chromosome 20q12-q13 in two clusters: centromeric and telomeric. This gene belongs to the telomeric cluster. Two alternatively spliced transcript variants have been found for this gene, and they encode distinct isoforms.
OTTHUMG00000046330	WFDC11	WAP four-disulfide core domain 11	This gene encodes a member of the WAP-type four-disulfide core (WFDC) domain

			family. The WFDC domain, or WAP signature motif, contains eight cysteines forming four disulfide bonds at the core of the protein, and functions as a protease inhibitor. Most WFDC gene members are localized to chromosome 20q12-q13 in two clusters: centromeric and telomeric. This gene belongs to the telomeric cluster.
OTTHUMG00000046412	WFDC12	WAP four-disulfide core domain 12	This gene encodes a member of the WAP-type four-disulfide core (WFDC) domain family. The WFDC domain, or WAP signature motif, contains eight cysteines forming four disulfide bonds at the core of the protein, and functions as a protease inhibitor. Most WFDC gene members are localized to chromosome 20q12-q13 in two clusters: centromeric and telomeric. This gene belongs to the centromeric cluster.
OTTHUMG00000046333	WFDC13	WAP four-disulfide core domain 13	This gene encodes a member of the WAP-type four-disulfide core (WFDC) domain family. The WFDC domain, or WAP signature motif, contains eight cysteines forming four disulfide bonds at the core of the protein, and functions as a protease inhibitor. Most WFDC gene members are localized to chromosome 20q12-q13 in two clusters: centromeric and telomeric. This gene belongs to the telomeric cluster.
OTTHUMG00000032594	WFDC2	WAP four-disulfide core domain 2	This gene generates multiple alternatively spliced transcript variants, which encode different protein isoforms. These isoforms contain one or two WAP-type four-disulfide core (WFDC) domains, and are thus members of the WFDC domain family. The WFDC domain, or WAP Signature motif, contains eight cysteines forming four disulfide bonds at the core of the protein, and functions as a protease inhibitor in many family members. This gene is expressed in pulmonary epithelial cells, and was also found to be expressed in some ovarian cancers. The encoded isoforms are small secretory proteins, which may be involved in sperm maturation.
OTTHUMG00000032614	WFDC3	WAP four-disulfide core domain 3	This gene encodes a member of the WAP-type four-disulfide core (WFDC) domain family. The WFDC domain, or WAP signature motif, contains eight cysteines forming four disulfide bonds at the core of the protein, and functions as a protease inhibitor. The encoded protein contains four WFDC domains. Most WFDC genes are localized to chromosome 20q12-q13 in two clusters: centromeric and telomeric. This gene belongs to the telomeric cluster. Several alternatively spliced transcript variants have been found for this gene, but the full-length nature of some variants has not been determined.
OTTHUMG00000046411	WFDC5	WAP four-disulfide core domain 5	This gene encodes a member of the WAP-type four-disulfide core (WFDC) domain family. Most WFDC proteins contain only one WFDC domain, and this encoded protein contains two WFDC domains. The WFDC domain, or WAP signature motif, contains eight cysteines forming four disulfide bonds at the core of the protein, and functions as a protease inhibitor. Most WFDC gene members are localized to chromosome 20q12-q13 in two clusters: centromeric and telomeric. This gene belongs to the centromeric cluster.
OTTHUMG00000046354	WFDC6	WAP four-disulfide core domain 6	This gene encodes a member of the WAP-type four-disulfide core (WFDC) domain family. The WFDC domain, or WAP signature motif, contains eight cysteines forming four disulfide bonds at the core of the protein, and functions as a protease inhibitor.

			Most WFDC gene members are localized to chromosome 20q12-q13 in two clusters: centromeric and telomeric. This gene belongs to the telomeric cluster.
OTTHUMG00000046342	WFDC8	WAP four-disulfide core domain 8	This gene encodes a member of the WAP-type four-disulfide core (WFDC) domain family. The WFDC domain, or WAP signature motif, contains eight cysteines forming four disulfide bonds at the core of the protein, and functions as a protease inhibitor. The encoded protein contains a Kunitz-inhibitor domain, in addition to three WFDC domains. Most WFDC genes are localized to chromosome 20q12-q13 in two clusters: centromeric and telomeric. This gene belongs to the telomeric cluster. Two alternatively spliced transcript variants have been found for this gene, and they encode the same protein.
OTTHUMG00000046332	WFDC9	WAP four-disulfide core domain 9	The WAP-type four-disulfide core (WFDC) domain, or WAP signature motif, contains eight cysteines forming four disulfide bonds at the core of the protein, and functions as a protease inhibitor in many members of the WFDC domain family. This gene encodes a protein which contains a WFDC domain, and is thus a member of the WFDC domain family. This gene and several other gene family members are clustered at 20q13.12.
OTTHUMG00000033071	WISP2	WNT1 inducible signaling pathway protein 2	This gene encodes a member of the WNT1 inducible signaling pathway (WISP) protein subfamily, which belongs to the connective tissue growth factor (CTGF) family. WNT1 is a member of a family of cysteine-rich, glycosylated signaling proteins that mediate diverse developmental processes. The CTGF family members are characterized by four conserved cysteine-rich domains: insulin-like growth factor-binding domain, von Willebrand factor type C module, thrombospondin domain and C-terminal cystine knot-like (CT) domain. The encoded protein lacks the CT domain which is implicated in dimerization and heparin binding. It is 72% identical to the mouse protein at the amino acid level. This gene may be downstream in the WNT1 signaling pathway that is relevant to malignant transformation. Its expression in colon tumors is reduced while the other two WISP members are overexpressed in colon tumors. It is expressed at high levels in bone tissue, and may play an important role in modulating bone turnover.
OTTHUMG00000032549	YWHAB	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, beta polypeptide	This gene encodes a protein belonging to the 14-3-3 family of proteins, members of which mediate signal transduction by binding to phosphoserine-containing proteins. This highly conserved protein family is found in both plants and mammals. The encoded protein has been shown to interact with RAF1 and CDC25 phosphatases, suggesting that it may play a role in linking mitogenic signaling and the cell cycle machinery. Two transcript variants, which encode the same protein, have been identified for this gene.
OTTHUMG00000032481	ZHX3	zinc fingers and homeoboxes 3	This gene encodes a member of the zinc fingers and homeoboxes (ZHX) gene family. The encoded protein contains two C2H2-type zinc fingers and five homeodomains and forms a dimer with itself or with zinc fingers and homeoboxes family member 1. In the nucleus, the dimerized protein interacts with the A subunit of the ubiquitous transcription factor nuclear factor-Y and may function as a transcriptional repressor.

OTTHUMG00000032709	ZNF313	zinc finger protein 313	
OTTHUMG00000032654	ZNF334	zinc finger protein 334	
OTTHUMG00000032637	ZNF335	zinc finger protein 335	The protein encoded by this gene enhances transcriptional activation by ligand-bound nuclear hormone receptors. However, it does this not by direct interaction with the receptor, but by direct interaction with the nuclear hormone receptor transcriptional coactivator NRC. The encoded protein may function by altering local chromatin structure.
OTTHUMG00000074023	ZSWIM1	zinc finger, SWIM-type containing 1	
OTTHUMG00000032627	ZSWIM3	zinc finger, SWIM-type containing 3	

Table A. 1 The list of 103 protein coding genes residing on 20q12-13.2 region. First column is the identification tag of the gene in VEGA Genome Browser. Sec column is the gene name set by HUGO gene nomenclature committee, third column is the long description of the gene name and the last column gives a summary about the function of the product of the gene.

<table>
<tr><th>External Gene ID</th><th>CpG Island Topology</th><th># of (CG)s</th><th>%GC</th><th>CpG island (Vega)</th><th>CpG island Searcher</th><th>Transcript Type</th></tr>
<tr><td>C20orf168-001</td><td>---</td><td>0</td><td>40.7</td><td>no</td><td>no</td><td>RT</td></tr>
<tr><td>WFDC11-001</td><td>---</td><td>0</td><td>40.7</td><td>no</td><td>no</td><td>RT</td></tr>
<tr><td>CHD6-002</td><td>---</td><td>0</td><td>41.3</td><td>no</td><td>no</td><td>AT</td></tr>
<tr><td>PRKCBP1-007</td><td>---</td><td>0</td><td>39.3</td><td>no</td><td>no</td><td>AT</td></tr>
<tr><td>PRKCBP1-009</td><td>---</td><td>0</td><td>41.3</td><td>no</td><td>no</td><td>AT</td></tr>
<tr><td>ZHX3-002</td><td>---</td><td>0</td><td>34.3</td><td>no</td><td>no</td><td>AT</td></tr>
<tr><td>C20orf10-001</td><td>--|---</td><td>1</td><td>49.0</td><td>no</td><td>no</td><td>RT</td></tr>
<tr><td>C20orf62-001</td><td>---|</td><td>1</td><td>45.7</td><td>no</td><td>no</td><td>RT</td></tr>
<tr><td>PI3-001</td><td>---|--------------------------------</td><td>1</td><td>45.3</td><td>no</td><td>no</td><td>RT</td></tr>
<tr><td>SEMG1-001</td><td>--|-------</td><td>1</td><td>41.7</td><td>no</td><td>no</td><td>RT</td></tr>
<tr><td>SPINT4-001</td><td>---|--</td><td>1</td><td>46.0</td><td>no</td><td>no</td><td>RT</td></tr>
<tr><td>WFDC10A-001</td><td>--|----</td><td>1</td><td>25.0</td><td>no</td><td>no</td><td>RT</td></tr>
<tr><td>WFDC10B-001</td><td>--|---</td><td>1</td><td>43.3</td><td>no</td><td>no</td><td>RT</td></tr>
<tr><td>SLC13A3-004</td><td>--|-----------------------------------</td><td>1</td><td>41.7</td><td>no</td><td>no</td><td>AT</td></tr>
<tr><td>SLC35C2-003</td><td>--|----</td><td>1</td><td>40.0</td><td>no</td><td>no</td><td>AT</td></tr>
<tr><td>UBE2C-003</td><td>---|--</td><td>1</td><td>44.7</td><td>no</td><td>no</td><td>AT</td></tr>
<tr><td>WFDC3-001</td><td>--------------------------------------|---</td><td>1</td><td>38.3</td><td>no</td><td>no</td><td>AT</td></tr>
<tr><td>ZHX3-004</td><td>--|--------------------------</td><td>1</td><td>41.0</td><td>no</td><td>no</td><td>AT</td></tr>
<tr><td>C20orf123-001</td><td>-----------------|-----------------|--</td><td>2</td><td>56.0</td><td>no</td><td>no</td><td>RT</td></tr>
<tr><td>C20orf157-001</td><td>--|---------------------------|------</td><td>2</td><td>36.3</td><td>no</td><td>no</td><td>RT</td></tr>
<tr><td>JPH2-001</td><td>--|---|----------------------</td><td>2</td><td>55.3</td><td>no</td><td>no</td><td>RT</td></tr>
<tr><td>MATN4-002</td><td>--|----------|-----</td><td>2</td><td>41.0</td><td>no</td><td>no</td><td>RT</td></tr>
<tr><td>R3HDML-001</td><td>--------------|------|--</td><td>2</td><td>59.3</td><td>no</td><td>no</td><td>RT</td></tr>
<tr><td>SPINLW1-001</td><td>------|---|-------------------</td><td>2</td><td>53.3</td><td>no</td><td>no</td><td>RT</td></tr>
<tr><td>WFDC6-002</td><td>--|----------------------|------------------------</td><td>2</td><td>55.7</td><td>no</td><td>no</td><td>RT</td></tr>
</table>

WFDC8-001	------|---|--	2	60.7	no	no	RT
C20orf119-007	----------------|---------------|--	2	47.3	no	no	AT
DNTTIP1-003	--|-------|----------	2	43.0	no	no	AT
GDAP1L1-002	----------------------------------|---|--	2	55.0	no	no	AT
L3MBTL-004	--|-|--------------------------	2	55.7	no	no	AT
PKIG-008	----------------------------------|-------------------------------|--------------------------------	2	46.7	no	no	AT
ZHX3-007	-------------|---|---	2	43.3	no	no	AT
SEMG2-001	-------------|--------|--------|---	3	41.7	no	no	RT
SLPI-001	--|-----------|--------------------------------|-----	3	49.0	no	no	RT
SPINT3-001	---|----------|----|---	3	55.0	no	no	RT
WFDC13-001	---|----------------|-----|--------------------	3	40.7	no	no	RT
WFDC9-001	--|--|---------------------|-----------------	3	53.7	no	no	RT
C20orf119-011	-----------------|------------------------------------|-------------|------------------------------	3	52.0	no	no	AT
DBNDD2-012	--------------|------|---|-------------------------	3	54.3	no	no	AT
ELMO2-005	--|-------------------------|---|------------	3	51.3	no	no	AT
PKIG-004	---------------------------|---|-----------------|-	3	49.0	no	no	AT
SGK2-002	-------------------------------|--------|-------------|--	3	56.3	no	no	AT
SPINLW1-002	----------|-------------------------------|---------------------------------|----------------------	3	34.0	no	no	AT
UBE2V1-007	----------|---|---|------------------	3	53.3	no	no	AT
C20orf165-001	-----------------------|---------------------------------|-------------------------|--------|------	4	48.7	no	no	RT
MMP9-001	----|--|------------|--|---------------------------	4	61.7	no	no	RT
WFDC5-002	----------------------------|--------------|---------------|------------|--------------------------	4	55.7	no	no	RT
C20orf119-004	--------------------|------------------------------|-|-------------------------------|-------------	4	47.7	no	no	AT
DBNDD2-010	--|-|----------||----------------------------------	4	53.3	no	no	AT
FAM112A-002	----|-----------------------------------|----------------------------|-----------------------------	4	47.3	no	no	AT
CHD6-008	----------|------|---|------------------|----------	4	44.3	no	no	AT
ELMO2-004	--|-----|-------|---|------------	4	41.3	no	no	AT
L3MBTL-011	-------------|--------------------------------|----------------------------|--------------------|--	4	54.3	no	no	AT

LPIN3-003	-------------------------|---------------|----------------------|---------|------------------------	4	57.3	no	no	AT
WFDC3-006	----|---------|--|---------------------|-----------------	4	50.7	no	no	AT
CDH22-001	-------|--------|--|--|---|	5	54.3	yes	yes	RT
SGK2-005	----------|--------|--|--------------|-----|-	5	55.0	no	no	RT
WFDC12-001	--|-----|-----|---|-|--	5	55.7	no	no	RT
C20orf119-005	-------------------------|--||--|-----------|----	5	43.0	no	no	AT
ELMO2-006	---------|------------------|-------|--|--|----------------	5	51.3	no	no	AT
ZNFX1-003	-|------------|------------|--|--------|-----------	5	48.7	no	no	AT
L3MBTL-010	---------------||-----|----------------|----------------|--	5	59.0	no	no	AT
PLCG1-004	------|------|----|--|-----|---------------------------	5	56.3	no	no	AT
PRKCBP1-006	-----|-----|--|--|----------------------|--------	5	48.3	no	no	AT
STAU1-008	|------||----------------|--------------|--	5	44.3	no	no	AT
SERINC3-002	-------------|--------------|-----------------|--||----------	5	46.3	no	no	AT
TNNC2-003	------||-----------------|-------------------------------|---------------------------|-------------	5	60.3	no	no	AT
ZHX3-008	---------|---------------------------------|-|------------------------------------|---------------|	5	59.0	no	no	AT
C20orf142-001	------------------|--|--------------------|---------------|-------------------------------|--------	6	43.0	yes	yes	RT
FAM112A-001	-----|---------|---------------|--|----------|--|--------	6	46.3	no	no	RT
C20orf119-009	---|-----||-|--------------||------	6	56.3	no	no	AT
PREX1-002	-------------------|---------------------------------|----|-----------------------|-|-------------|	6	58.7	no	no	AT
EYA2-004	----|-------|----------|----------||------------------------|---------------|----------------------	7	53.3	no	no	AT
SLC35C2-007	------------------------------------||-------|----------|--------|----------|-------------|--------	7	56.0	no	no	AT
SLC13A3-003	---------------------|--------------|----|---------------------------|-|-------------|--------|----	8	61.3	no	no	AT
ZSWIM3-001	-------|------|-----|---------------|------------------|-------|---|--|-----------------------|----	9	52.3	no	yes	RT
DNTTIP1-005	---------------------------------|--|-|---------|--------|-----|--------------------|----|----|----	9	52.3	no	yes	AT
L3MBTL-005	--------------|-|--------------|----------------|---------------|-----|---------|-----------|----|-	9	66.3	no	no	RT
ELMO2-007	--------------------------|--|---|----||------|---------------------------------|----|------|------	9	51.7	no	no	AT
SULF2-003	----|-------------------------|--------------------------|--------------|---|----|----|----|-|-----	9	59.7	no	no	AT
PRKCBP1-003	--------------------|-----|--------||----|-|----|------|----|---------------------------------|----	10	48.3	no	no	RT

<table>
<tr><td>DBNDD2-006</td><td>-------|-|-------|-----------|--|------------------||----|-----|-------------------------------|---</td><td>10</td><td>61.3</td><td>no</td><td>no</td><td>AT</td></tr>
<tr><td>HNF4A-003</td><td>-----------|------------|--------------|----||--|--------|--------------------|---------------|--|-</td><td>10</td><td>60.3</td><td>no</td><td>no</td><td>AT</td></tr>
<tr><td>SULF2-005</td><td>|--|----|------------------------|---|-------|-------|--------|---------------------------|---|----</td><td>10</td><td>60.7</td><td>no</td><td>no</td><td>AT</td></tr>
<tr><td>SGK2-007</td><td>-----------------|-|-------------------|--|------||-|----------|-----------------------|-----------</td><td>10</td><td>63.0</td><td>no</td><td>no</td><td>AT</td></tr>
<tr><td>SLC35C2-006</td><td>----|----------------------------|------------|---------|-------------|-----|---|-|-----|-------|--</td><td>10</td><td>60.3</td><td>no</td><td>no</td><td>AT</td></tr>
<tr><td>UBE2V1-004</td><td>---|----------------------|--------|----------------------------------|--|------|---|-|--|--|------</td><td>10</td><td>59.3</td><td>no</td><td>no</td><td>AT</td></tr>
<tr><td>HNF4A-001</td><td>---|---------------------------|-------|-----------|----------|-------------|-|---|--|----|--------</td><td>11</td><td>60.0</td><td>no</td><td>no</td><td>RT</td></tr>
<tr><td>C20orf100-001</td><td>--------------------------|---------|-----------|-----------|--||----------|---|--|----|-|---------</td><td>11</td><td>55.0</td><td>no</td><td>no</td><td>AT</td></tr>
<tr><td>C20orf100-004</td><td>------------------------||-------------|------|----------------|---|------|--------|----------|----</td><td>11</td><td>53.7</td><td>no</td><td>no</td><td>AT</td></tr>
<tr><td>JPH2-002</td><td>---|-----------------------------|----------|--||--|--|---------|--|----||---|---------------------</td><td>12</td><td>64.0</td><td>no</td><td>no</td><td>AT</td></tr>
<tr><td>SLC2A10-001</td><td>|---|------------||-----------------------|---------------------|---|--|----|----|---------||-|---</td><td>13</td><td>66.7</td><td>yes</td><td>yes</td><td>RT</td></tr>
<tr><td>ACOT8-001</td><td>----|||-||--------|------|------|--------------|------------------|--------|--|--|-----------------</td><td>13</td><td>54.2</td><td>no</td><td>yes</td><td>AT</td></tr>
<tr><td>WISP2-002</td><td>---||-------------|----|--------|-|--|-----|---------------|---------------------|---||------|-----</td><td>13</td><td>63.3</td><td>no</td><td>no</td><td>RT</td></tr>
<tr><td>TP53RK-001</td><td>--------------------|--------|------------------------------|--------||||--|--||--|--|---------|-|-</td><td>14</td><td>49.7</td><td>yes</td><td>yes</td><td>RT</td></tr>
<tr><td>WISP2-001</td><td>--------------------------------||-------|--------------------------|--------------------|||||||--|</td><td>14</td><td>57.7</td><td>no</td><td>no</td><td>AT</td></tr>
<tr><td>SERINC3-001</td><td>----------------|---|---|-||||-|--||-|-||----|-|----|---</td><td>16</td><td>54.0</td><td>yes</td><td>yes</td><td>RT</td></tr>
<tr><td>C20orf111-002</td><td>|------|-----||----|------|--|------|-|---------|-------|------------|---||----------------|-|-----</td><td>16</td><td>59.0</td><td>yes</td><td>yes</td><td>AT</td></tr>
<tr><td>CD40-001</td><td>------|-|---------------------|----|----|-------------|------|--|---|------|----|------------|||---</td><td>16</td><td>64.0</td><td>no</td><td>no</td><td>RT</td></tr>
<tr><td>CHD6-006</td><td>||--||----||-----|----|---|------------|--|---------||---||---|-|-----------------------------------</td><td>17</td><td>56.0</td><td>yes</td><td>yes</td><td>RT</td></tr>
<tr><td>EYA2-005</td><td>----------------------------|---------|-||----------------|----|--|------------||--||--|-----||||-|</td><td>17</td><td>60.7</td><td>yes</td><td>yes</td><td>RT</td></tr>
<tr><td>PLTP-002</td><td>---------------|--|||-|--||--|-|-----|---------|--|-------|--|----------|-----|------------------|-</td><td>17</td><td>61.0</td><td>yes</td><td>yes</td><td>AT</td></tr>
<tr><td>UBE2C-004</td><td>-|---||-------------------||----|--|--------|----------|-|--|-----||---|---------|-----|-------|---</td><td>17</td><td>65.0</td><td>YES</td><td>yes</td><td>AT</td></tr>
<tr><td>GDAP1L1-001</td><td>--|--------------------|----|-------------|---||---------|---|------|-----|-|--------||------|---||</td><td>17</td><td>66.3</td><td>yes</td><td>no</td><td>RT</td></tr>
<tr><td>DDX27-001</td><td>----------------------------------|--|-----------|---|-|---|----|-|----||--|-----|--|---|--||---|-</td><td>17</td><td>49.3</td><td>no</td><td>yes</td><td>RT</td></tr>
<tr><td>KCNS1-001</td><td>-----------------------------|---|-------|-|-|-----------------|-----||-||-|-----|-|-||---||------</td><td>17</td><td>59.0</td><td>no</td><td>yes</td><td>RT</td></tr>
<tr><td>ZSWIM1-002</td><td>-|------|--------------------------|-||||-|-|--|------|------|--|--|-----------|--------||---------</td><td>17</td><td>57.7</td><td>no</td><td>no</td><td>RT</td></tr>
<tr><td>PKIG-002</td><td>-----------------------||----------------|------------|---|-----|---|----|---|-----||-|--||-|----||</td><td>18</td><td>60.3</td><td>yes</td><td>yes</td><td>RT</td></tr>
<tr><td>KCNK15-001</td><td>------------------|---|----|-------|------|----------------|--|---|---|--|--|-|------|-|-|-||----|-</td><td>19</td><td>65.3</td><td>yes</td><td>yes</td><td>RT</td></tr>
</table>

<table>
<tr><td>WFDC2-002</td><td>---------------------|---|-|---|----------|------------------|-|--|--------||-----|---|||---|----|-</td><td>19</td><td>66.3</td><td>yes</td><td>no</td><td>RT</td></tr>
<tr><td>LPIN3-001</td><td>----|------|----|------|-|-|---|--|-|---|----------------|--|-----|--------|-|-----|---------|---||</td><td>19</td><td>69.0</td><td>no</td><td>no</td><td>RT</td></tr>
<tr><td>CSE1L-001</td><td>-------------||---------|-|-----------------|----|-|---|----------|---|-||--|-----|-|---|-----|-||-</td><td>20</td><td>61.7</td><td>yes</td><td>yes</td><td>RT</td></tr>
<tr><td>PLTP-001</td><td>-------|------------------||---|---------|--|-||------------------|--|---|------|--|||-----|---|-||</td><td>20</td><td>69.3</td><td>yes</td><td>yes</td><td>RT</td></tr>
<tr><td>SULF2-001</td><td>-||--------|------------|----|--|---|----|---|--|------------|--|-------|-------------|---|---|||--</td><td>20</td><td>59.3</td><td>yes</td><td>yes</td><td>RT</td></tr>
<tr><td>TOMM34-001</td><td>----------------------|------------------|------|--||-----------|-|--------|-||--||-|----|--|--||-|</td><td>20</td><td>58.7</td><td>yes</td><td>yes</td><td>RT</td></tr>
<tr><td>UBE2V1-009</td><td>--------|---------------|-|---|-||-------|--|-------|--|-|||--------|-|--------|--|-----|--|-|-----</td><td>20</td><td>54.3</td><td>no</td><td>no</td><td>AT</td></tr>
<tr><td>C20orf119-017</td><td>------------------------|---|---------------|---------|---||-|-|--|--||-|----|-||-----|-|-||-----|-</td><td>21</td><td>69.0</td><td>yes</td><td>yes</td><td>RT</td></tr>
<tr><td>IFT52-001</td><td>-------|-------|------|--|-||--|-|--|-||-|--||---|--|||||--</td><td>21</td><td>54.3</td><td>yes</td><td>yes</td><td>RT</td></tr>
<tr><td>PTGIS-001</td><td>--|------------------------------||-|--|-|----------|--|----|---|||--|---------|--|--|--|--||---||-</td><td>21</td><td>75.3</td><td>yes</td><td>yes</td><td>RT</td></tr>
<tr><td>ACOT8-004</td><td>-----|-----|------|--------||-||-|----------|-|----------||||-----|-|--|-|---||--|-----|-----------</td><td>22</td><td>63.3</td><td>no</td><td>yes</td><td>RT</td></tr>
<tr><td>ZNF334-001</td><td>|-|------|-------------------------|-|||------|-||--|---|--|-|---|---------|--|-|---|------|----||-</td><td>22</td><td>60.0</td><td>no</td><td>yes</td><td>RT</td></tr>
<tr><td>SLC13A3-001</td><td>-------------|---|----------|-|----|---|-------------|---|---|-------|||--|---|-||-|-|-|--||---|---</td><td>23</td><td>69.0</td><td>no</td><td>yes</td><td>RT</td></tr>
<tr><td>SLC12A5-001</td><td>----|----|-------|-----------|---|---------|-----|----|---|-----||----|||--|--|-|---|||-|------|-|-</td><td>24</td><td>68.0</td><td>yes</td><td>yes</td><td>RT</td></tr>
<tr><td>ZNF335-001</td><td>-|-----|--||---------------|||---------|-----|-|--|-------|--|----|----|---|-|-----|-||-|----|--|--</td><td>24</td><td>56.7</td><td>yes</td><td>yes</td><td>RT</td></tr>
<tr><td>PIGT-003</td><td>-----|----|------|-----|---------------||-|-||---|--|-|-|-|-----|-----|----|-||--|---|--||--|----|-</td><td>25</td><td>61.7</td><td>yes</td><td>yes</td><td>RT</td></tr>
<tr><td>YWHAB-002</td><td>|------|---------|-|------|--------|--|--|------|-------|-|--|-|---|-|------|--|-|---||--|-||----|-</td><td>25</td><td>69.7</td><td>yes</td><td>yes</td><td>RT</td></tr>
<tr><td>GDAP1L1-006</td><td>--||------|---|||---------|----|---------------||--|-|---|--||-|-|--|------------------|----||--|-|</td><td>25</td><td>70.0</td><td>yes</td><td>no</td><td>AT</td></tr>
<tr><td>C20orf121-002</td><td>--------------------------|---|------------|-|--|--------|||-||--||--|--|--|--|--|||-|--|----||--||</td><td>25</td><td>65.3</td><td>no</td><td>yes</td><td>RT</td></tr>
<tr><td>ELMO2-003</td><td>---------|-|-||----|---|-----------------|||-|----|-|--|-|----|-||----||-|-|--------|------|---|||-</td><td>26</td><td>68.7</td><td>yes</td><td>yes</td><td>RT</td></tr>
<tr><td>UBE2C-001</td><td>--||-|---------|----|--|---||--------|---|-||--------|-----||-------||||-|------|--|------||-|-----</td><td>26</td><td>59.0</td><td>yes</td><td>yes</td><td>RT</td></tr>
<tr><td>TP53RK-002</td><td>---|-||-|---|-|--|---------------|----|----|---|--|-|-|------|------||-|--|-------|----||----||----</td><td>26</td><td>59.7</td><td>yes</td><td>yes</td><td>AT</td></tr>
<tr><td>C20orf67-001</td><td>|-|----||-||--|--|--|-----|-|----------|----------|------|-------|-----|---|------||||||----|--|-</td><td>27</td><td>67.0</td><td>yes</td><td>yes</td><td>RT</td></tr>
<tr><td>PPGB-002</td><td>----|---------|----||------|--||---|-----|||----||------||--|-|-------|-||||---|---------------|-|-</td><td>27</td><td>66.7</td><td>yes</td><td>yes</td><td>RT</td></tr>
<tr><td>RBPSUHL-003</td><td>|--|-|------|-|-||--|--|-----||--|-|----|---------|--|----|-|------------|--|---------|-----|--||--</td><td>27</td><td>70.0</td><td>yes</td><td>yes</td><td>RT</td></tr>
<tr><td>SNX21-006</td><td>-|----|||----|-|----|-|||---|------|---|----------|-------|--|--------|----|-|-|-----||------|-|--|</td><td>28</td><td>73.0</td><td>yes</td><td>yes</td><td>AT</td></tr>
<tr><td>PPGB-005</td><td>---|---|---||-----|----|------|||-|-|-|--|---|---|-|---|------|----|-||--|-|-------||------|-------</td><td>28</td><td>66.8</td><td>yes</td><td>yes</td><td>AT</td></tr>
<tr><td>UBE2V1-014</td><td>----------|-|----------|-|----------------|-|---||-|-|||-|-|||-|--|--|--------|----||------|-|-|-||</td><td>28</td><td>66.7</td><td>yes</td><td>yes</td><td>AT</td></tr>
</table>

DBNDD2-005	----||----|---|--|-|-|----|---|-|----|------|||--||-|-------||-|||--|--------||------------||-------|	29	72.7	yes	yes	RT
KCNB1-001	----------|-|-|--------||---||---|--|-----|--|-----|---------|-------|---|||-|---|-||------||-||||-	29	72.7	yes	yes	RT
PPGB-001	----|--|||---||-|---||--||----||---|--|--|-----|||------|-|-|------|-------||----||-----|-|--------	29	69.7	yes	yes	AT
NCOA5-002	---------|---|-|-------|-------|---|-|---|-------||-----|--||--|--||-|-|||-------|-||---|-||-|--|--	30	66.7	yes	yes	RT
NEURL2-001	--|-|----|----||-|---|--|----||---||-|-|-|-----|--------|------||-------|-|--||-|----|------|---|-|	30	72.0	yes	yes	RT
SNX21-003	--||--|-----|-------|------------|--||-||-||---|-||----|||--|---||----|-|----|----|---|||----|--|--	31	75.7	yes	yes	RT
SLC35C2-005	----|--------|----|-----------|----||-|---||--|-|-|------||--|-|||-|||-|-|--|---|------|-----||--||	31	69.0	yes	yes	RT
SLC9A8-001	-|-|------|---|--------|--||--|---------|-----|---|----|-||||--|-||-|||--|-|---|----||-|-|-----|---	31	69.3	yes	yes	RT
SPATA2-001	----|-----||--|-----|-||-|-|-|--|---|-----||-|--||-|--||-|-----|-|------|-|--|-|---|---------||----	31	66.7	yes	yes	RT
NCOA3-001	-----------|-|--|----|--|-|-||--|||--||------|----|||-----|---|--|-------|-||-------|--||-|-|--|-|-	32	69.0	yes	yes	RT
C20orf111-001	-|---|------|------|-|-|----|-|-|---|-||-|-|-||-||-|---------------|-|-|-|--|||-|------|--|--|----|	33	68.3	yes	yes	RT
ZNFX1-004	----|-|-||--|-------------|--|---||---|-|||||------|----||------||--|-----||--|---|-|--||--|--|||--	33	63.3	yes	yes	RT
SDC4-001	-------------|---------||-|--|-------|||-|--|--|-----|-|-|||-|-|-||-|--|-----||||---||--|----||----	33	74.0	yes	yes	RT
TNNC2-001	|------|---|||-|----|-|----|||----|---|-----||-----||--|--||-|----|||---|-|||-||--|------------|---	33	76.0	yes	yes	RT
ZNF313-001	--|--|-----|--|--|-----------|-----|---|-|----|-|--|--|---|-|-|--|-|----|-|||||-|----||--|-|---|--|	33	69.0	yes	yes	RT
STAU1-002	----||--------|--|---|-|-|--|--|--|---||--||--||---|----||-----|-|---|-|-|-||----|----|-||--|---|-|	34	76.7	yes	yes	RT
C20orf100-002	--|---|-|-|----|--------------||--|||-------------||-|-|-||-||----||-|--|-|---||---|---|-|--|-|-|--	34	68.3	yes	yes	AT
TOP1-001	-----||---|------|-|-||-|--|-----|-|--|-|-|-------|-|-||||----|||-|----|------------|||-|---||||----	35	66.3	yes	yes	RT
ZNFX1-002	---||-|||-----|--|-|-|||-----|-------|-------|-||-|---|--|-|--|||-|--|---|----|||----|||-|---|-----	35	78.0	yes	yes	AT
STK4-003	----------|-|---|---|-|-|---|-------|-|---|----|-|-||-||----||-|---|--|-|-|||||-|--|-|------||----|	35	69.7	no	yes	RT
DBNDD2-008	|-||-|-|----|||---||||-|------|--|------|----|------|||----|-|||-|-|--|-|-----|-|---|-|---||--|----	36	73.3	yes	yes	AT
DNTTIP1-006	--|-|--|-----||-|-----|--|-|---||-|||-----|-|----|-----||--||-|---|--|||-|||---|-----||||-|--------	36	72.3	no	yes	RT
WFDC3-002	---------|-||-|-----|---||-|-||-||---|||---||----|----|-|-----|||--||--|-|--|------||-|-----|--|-|-	36	72.1	no	yes	RT
SFRS6-002	-------|-------------|--|-||--|-|-------||||-||-||--|---|--||---|---||||------||-----||-|--|-|-|-||	37	66.0	yes	yes	RT
SNAI1-001	|-|-|-|-|---||---|||||-|--||---|---||-|--|-||-----|----|--------|------||-----||||--|----|---|-||--	37	72.3	yes	yes	RT
ZHX3-006	-|--|---------|-||----|-|||---|-||--|-|-|--|--|------|--|-|--|--|---|--||-|-|-|--||-|-|----|-||---||	38	77.7	yes	yes	RT
MAFB-001	-----|--|--||---||-|-||-|---||-|-----|-|-||--||-|----||-||-|-|-------|---||||------||-|-|-||--|----	39	76.3	yes	yes	RT
EMILIN3-001	|-||-|-|------|-----|||-|-------------|-|--|||-|--|---||||||------|---|-|---||--|--||--|||-|||-|-||	40	77.3	yes	yes	RT

MYBL2-001	---|-|-||----|--------|-|--||-|---|-------||-||---|||||-|--------||||----|-|-||----|-|-||||--||-|--	41	77.7	yes	yes	RT
EYA2-003	-||--|-----||||-|-----|--|-||---|||--|||||--||-||||--||-----|-------|--||---||----|--|-|----|-||--|	42	79.7	yes	yes	AT
ADA-001	|-----|-------|---|--||--|||----||-||-|--|-|||-||-------||--|||-|--||--|-|-||||-|-|--||---|-|-|--|-	44	78.0	yes	yes	RT
ARFGEF2-001	---|-|---|-|||--||-|-|--|-|-||---|||--|-|---|||||-|-|-||-||-|---|-|-||-|||-|--||-|-----------|-|---	45	76.7	yes	yes	RT
SULF2-002	-||-|||---|--||-|-|-|-||-------|-|-|--|---||-|-|-|----||||---|-|-|||--|---------|||----|-||||||-|-|	45	82.7	yes	yes	AT
C20orf100-003	-----||--|||--|-----|-||--|--|---||-|---||||-|-|-|-|----|-||--|-|-|-|-----|||----||-||||||||--||---	46	84.0	yes	yes	RT
PLCG1-005	|-||-||-|-|-|----|||-|||-|||----|-|||---|||-|-------||||---||-|----|||-||--|--------||||---||----||	46	81.7	yes	yes	RT
B4GALT5-001	||-------|---|-|------|--|---|----|||--||||-||---||-||---|--|||-|-|-||-|||-|---|-|---|-|||||--||-||	48	77.7	yes	yes	RT
PTPRT-004	||||---|--|-|--|-|-|---|--|||||---|-||-|--||---||---||||||---|||---|-|-||-----|||-|-|-||||---------	48	81.7	yes	yes	RT
PREX1-001	-----|--||-|---|---|-|---|-||--||--|-|--||----|||||-|-||--|-|||-||---|-|---|-|||---||||||----|-|-|-	50	84.7	yes	yes	RT
UBE2V1-005	||||------|---|------|-|-|----|---|--|---|-||---|-||-|--||-|||-|---|||||--|-|||||||--|--||--|||-|--	51	79.3	yes	yes	RT
RIMS4-002	--|-|-||--||---|-||-|||-||||||----|-|-||||----|||-||||--|-|-|-|-|-||-|||||-||---|-----|-|||||||-||	62	86.7	yes	yes	RT

Table A. 2 CpG content of 177 promoter regions. 250 bp upstream and 50 bp downstream of the TSS is taken as the promoter sequence. Each “CG” dinucleotide is shown with “|”, other dinucleotides are shown with “-“. If there are no 3 consecutive “CG” dinucleotides in the sequence, it is represented as “-“ symbol, all other combinations are represented “|” for better display purposes. The grey boxes represent Vega CpG islands rejected by a more stringent definition (see 3.2.1). Yellow boxes represent new CpG islands meeting the conditions of the new definition but not detected by Vega. Green boxes represent CpG islands extending to the first exon of the gene.

External Gene ID	Transcript Type	CpG island Searcher	FirstEF	Eponine
ZHX3-006	RT	yes	yes	yes
SFRS6-002	RT	yes	yes	yes
C20orf100-003	RT	yes	yes	yes
DNTTIP1-006	RT	yes	yes	yes
WFDC3-002	RT	yes	yes	yes
PREX1-001	RT	yes	yes	yes
ARFGEF2-001	RT	yes	yes	yes
PTPRT-004	RT	yes	yes	yes
PLTP-001	RT	yes	yes	yes
SNAI1-001	RT	yes	yes	yes
MAFB-001	RT	yes	yes	yes
ELMO2-003	RT	yes	yes	yes
KCNS1-001	RT	yes	yes	yes
PLCG1-005	RT	yes	yes	yes
SDC4-001	RT	yes	yes	yes
UBE2V1-005	RT	yes	yes	yes
RIMS4-002	RT	yes	yes	yes
EMILIN3-001	RT	yes	yes	yes
IFT52-001	RT	yes	yes	no
C20orf111-001	RT	yes	yes	no
YWHAB-002	RT	yes	yes	no
C20orf119-017	RT	yes	yes	no
PIGT-003	RT	yes	yes	no
DBNDD2-005	RT	yes	yes	no
TNNC2-001	RT	yes	yes	no
SNX21-003	RT	yes	yes	no
NEURL2-001	RT	yes	yes	no
ZSWIM3-001	RT	yes	yes	no
C20orf67-001	RT	yes	yes	no
ZNF335-001	RT	yes	yes	no
SLC12A5-001	RT	yes	yes	no
ZNF334-001	RT	yes	yes	no
SULF2-001	RT	yes	yes	no
SLC13A3-001	RT	yes	yes	no

External Gene ID	Tanscript Type	CpG island Searcher	FirstEF	Eponine
DBNDD2-008	AT	yes	yes	yes
SULF2-002	AT	yes	yes	yes
ZNFX1-002	AT	yes	yes	yes
UBE2C-004	AT	yes	yes	yes
EYA2-003	AT	yes	yes	yes
C20orf100-002	AT	yes	yes	no
DNTTIP1-005	AT	yes	yes	no
SNX21-006	AT	yes	yes	no
PLTP-002	AT	yes	yes	no
PPGB-005	AT	yes	yes	no
TP53RK-002	AT	yes	yes	no
PPGB-001	AT	yes	no	yes
UBE2V1-014	AT	yes	no	yes
C20orf111-002	AT	yes	no	no
ACOT8-004	AT	yes	no	no
GDAP1L1-006	AT	no	yes	no
HNF4A-003	AT	no	yes	no
ZHX3-002	AT	no	no	no
ZHX3-004	AT	no	no	no
ZHX3-007	AT	no	no	no
ZHX3-008	AT	no	no	no
CHD6-002	AT	no	no	no
CHD6-008	AT	no	no	no
L3MBTL-004	AT	no	no	no
L3MBTL-010	AT	no	no	no
L3MBTL-011	AT	no	no	no
FAM112A-002	AT	no	no	no
C20orf100-001	AT	no	no	no
C20orf100-004	AT	no	no	no
GDAP1L1-002	AT	no	no	no
C20orf119-005	AT	no	no	no
C20orf119-007	AT	no	no	no
C20orf119-009	AT	no	no	no

ACOT8-001	RT	yes	yes	no
CSE1L-001	RT	yes	yes	no
SLC35C2-005	RT	yes	yes	no
KCNB1-001	RT	yes	yes	no
RBPSUHL-003	RT	yes	yes	no
STK4-003	RT	yes	yes	no
NCOA3-001	RT	yes	yes	no
PKIG-002	RT	yes	yes	no
DDX27-001	RT	yes	yes	no
CDH22-001	RT	yes	yes	no
PTGIS-001	RT	yes	yes	no
SERINC3-001	RT	yes	yes	no
ZNFX1-004	RT	yes	yes	no
SLC9A8-001	RT	yes	yes	no
TP53RK-001	RT	yes	yes	no
TOMM34-001	RT	yes	no	yes
NCOA5-002	RT	yes	no	yes
TOP1-001	RT	yes	no	yes
MYBL2-001	RT	yes	no	yes
ADA-001	RT	yes	no	yes
B4GALT5-001	RT	yes	no	yes
C20orf121-002	RT	yes	no	yes
CHD6-006	RT	yes	no	no
C20orf142-001	RT	yes	no	no
KCNK15-001	RT	yes	no	no
SLC2A10-001	RT	yes	no	no
STAU1-002	RT	yes	no	no
SPATA2-001	RT	yes	no	no
ZNF313-001	RT	yes	no	no
UBE2C-001	RT	yes	no	no
EYA2-005	RT	yes	no	no
PPGB-002	RT	yes	no	no
L3MBTL-005	RT	no	yes	no
HNF4A-001	RT	no	yes	no
WFDC2-002	RT	no	yes	no
C20orf119-011	AT	no	no	no

DBNDD2-006	AT	no	no	no
DBNDD2-010	AT	no	no	no
DBNDD2-012	AT	no	no	no
SPINLW1-002	AT	no	no	no
DNTTIP1-003	AT	no	no	no
WFDC3-001	AT	no	no	no
WFDC3-006	AT	no	no	no
TNNC2-003	AT	no	no	no
PRKCBP1-006	AT	no	no	no
PRKCBP1-007	AT	no	no	no
PRKCBP1-009	AT	no	no	no
SULF2-003	AT	no	no	no
SULF2-005	AT	no	no	no
PREX1-002	AT	no	no	no
STAU1-008	AT	no	no	no
ZNFX1-003	AT	no	no	no
JPH2-002	AT	no	no	no
UBE2C-003	AT	no	no	no
EYA2-004	AT	no	no	no
SLC13A3-003	AT	no	no	no
SLC13A3-004	AT	no	no	no
SLC35C2-003	AT	no	no	no
SLC35C2-006	AT	no	no	no
SLC35C2-007	AT	no	no	no
SGK2-002	AT	no	no	no
SGK2-007	AT	no	no	no
LPIN3-003	AT	no	no	no
ELMO2-004	AT	no	no	no
ELMO2-005	AT	no	no	no
ELMO2-006	AT	no	no	no
ELMO2-007	AT	no	no	no
WISP2-001	AT	no	no	no
PLCG1-004	AT	no	no	no
UBE2V1-004	AT	no	no	no
UBE2V1-007	AT	no	no	no
JPH2-001	RT	no	yes	no

CD40-001	RT	no	yes	no
LPIN3-001	RT	no	yes	no
FAM112A-001	RT	no	no	no
R3HDML-001	RT	no	no	no
GDAP1L1-001	RT	no	no	no
C20orf62-001	RT	no	no	no
SEMG1-001	RT	no	no	no
SEMG2-001	RT	no	no	no
PI3-001	RT	no	no	no
C20orf10-001	RT	no	no	no
SPINT3-001	RT	no	no	no
SPINLW1-001	RT	no	no	no
SPINT4-001	RT	no	no	no
C20orf168-001	RT	no	no	no
C20orf165-001	RT	no	no	no
C20orf157-001	RT	no	no	no
C20orf123-001	RT	no	no	no
PRKCBP1-003	RT	no	no	no
MATN4-002	RT	no	no	no
MMP9-001	RT	no	no	no
SGK2-005	RT	no	no	no
SLPI-001	RT	no	no	no
WFDC11-001	RT	no	no	no
WFDC10A-001	RT	no	no	no
WFDC9-001	RT	no	no	no
WFDC13-001	RT	no	no	no
WFDC10B-001	RT	no	no	no
WFDC8-001	RT	no	no	no
WFDC6-002	RT	no	no	no
WFDC5-002	RT	no	no	no
WFDC12-001	RT	no	no	no
WISP2-002	RT	no	no	no
ZSWIM1-002	RT	no	no	no

Table A. 3 The prediction profiles of 103 representative transcripts (RT)

UBE2V1-009	AT	no	no	no
SERINC3-002	AT	no	no	no
PKIG-004	AT	no	no	no
PKIG-008	AT	no	no	no
C20orf119-004	AT	no	no	no

Table A. 4 Prediction profiles of alternative transcripts (AT)

Expression in HeLa S3	Expression in NTERA-D1	Vega Gene ID	HUGO Gene Name	Transcript count
P	P	OTTHUMG00000033045	ACOT8	3
P	P	OTTHUMG00000033081	ADA	3
P	P	OTTHUMG00000032687	ARFGEF2	1
P	P	OTTHUMG00000033086	B4GALT5	1
P	P	OTTHUMG00000032582	C20orf10	1
A	P	OTTHUMG00000032517	C20orf100	4
P	P	OTTHUMG00000032518	C20orf111	2
P	P	OTTHUMG00000032553	C20orf119	8
P	P	OTTHUMG00000032536	C20orf121	3
NO PROBE	NO PROBE	OTTHUMG00000032652	C20orf123	1
P	P	OTTHUMG00000032522	C20orf142	1
NO PROBE	NO PROBE	OTTHUMG00000032647	C20orf157	1
A	A	OTTHUMG00000032628	C20orf165	1
NO PROBE	NO PROBE	OTTHUMG00000032611	C20orf168	1
A	A	OTTHUMG00000032533	C20orf62	2
P	P	OTTHUMG00000032635	C20orf67	1
P	P	OTTHUMG00000033053	CD40	3
A	P	OTTHUMG00000033073	CDH22	1
P	P	OTTHUMG00000032487	CHD6	4
P	P	OTTHUMG00000033046	CSE1L	1
P	P	OTTHUMG00000032576	DBNDD2	13
P	P	OTTHUMG00000033072	DDX27	1
P	P	OTTHUMG00000032610	DNTTIP1	5
P	P	OTTHUMG00000033070	ELMO2	8
P	P	OTTHUMG00000046304	EMILIN3	1
A	A	OTTHUMG00000033041	EYA2	3
A	A	OTTHUMG00000032510	FAM112A	2
A	P	OTTHUMG00000032530	GDAP1L1	5
A	A	OTTHUMG00000032531	HNF4A	4
P	P	OTTHUMG00000032513	IFT52	2
A	A	OTTHUMG00000033037	JPH2	2
A	A	OTTHUMG00000033051	KCNB1	1

P	A	OTTHUMG00000032544	KCNK15	1
A	A	OTTHUMG00000033079	KCNS1	1
P	P	OTTHUMG00000032503	L3MBTL	13
A	A	OTTHUMG00000033056	LPIN3	2
A	P	OTTHUMG00000033052	MAFB	1
A	P	OTTHUMG00000033043	MATN4	3
A	P	OTTHUMG00000033044	MMP9	1
P	P	OTTHUMG00000033062	MYBL2	1
P	P	OTTHUMG00000033061	NCOA3	3
P	P	OTTHUMG00000032639	NCOA5	2
A	A	OTTHUMG00000032626	NEURL2	1
P	A	OTTHUMG00000032567	PI3	1
P	P	OTTHUMG00000032574	PIGT	3
P	P	OTTHUMG00000033065	PKIG	7
P	P	OTTHUMG00000033082	PLCG1	2
A	P	OTTHUMG00000033047	PLTP	2
P	P	OTTHUMG00000033078	PPGB	5
A	P	OTTHUMG00000032685	PREX1	2
P	P	OTTHUMG00000032667	PRKCBP1	5
A	P	OTTHUMG00000033077	PTGIS	1
A	A	OTTHUMG00000033040	PTPRT	7
A	A	OTTHUMG00000032524	R3HDML	1
A	P	OTTHUMG00000033055	RBPSUHL	3
A	A	OTTHUMG00000032546	RIMS4	1
P	P	OTTHUMG00000033083	SDC4	1
A	P	OTTHUMG00000032565	SEMG1	1
A	A	OTTHUMG00000032566	SEMG2	1
P	P	OTTHUMG00000033087	SERINC3	2
P	P	OTTHUMG00000032502	SFRS6	2
A	A	OTTHUMG00000033054	SGK2	6
A	A	OTTHUMG00000032638	SLC12A5	2
A	P	OTTHUMG00000033042	SLC13A3	4
A	P	OTTHUMG00000032657	SLC2A10	1
P	P	OTTHUMG00000033050	SLC35C2	7

P	P	OTTHUMG00000032710	SLC9A8	1
P	A	OTTHUMG00000033075	SLPI	1
A	P	OTTHUMG00000033048	SNAI1	1
P	P	OTTHUMG00000032624	SNX21	12
P	P	OTTHUMG00000032704	SPATA2	1
P	A	OTTHUMG00000032588	SPINLW1	5
A	A	OTTHUMG00000032585	SPINT3	1
NO PROBE	NO PROBE	OTTHUMG00000032604	SPINT4	1
P	P	OTTHUMG00000032691	STAU1	8
P	P	OTTHUMG00000033059	STK4	2
A	P	OTTHUMG00000032675	SULF2	4
A	A	OTTHUMG00000032623	TNNC2	2
P	P	OTTHUMG00000032552	TOMM34	1
P	P	OTTHUMG00000033057	TOP1	1
P	P	OTTHUMG00000085887	TP53RK	2
P	P	OTTHUMG00000033038	UBE2C	3
P	P	OTTHUMG00000033085	UBE2V1	9
A	A	OTTHUMG00000046331	WFDC10A	1
A	A	OTTHUMG00000046334	WFDC10B	1
A	A	OTTHUMG00000046330	WFDC11	2
A	A	OTTHUMG00000046412	WFDC12	1
A	A	OTTHUMG00000046333	WFDC13	1
P	P	OTTHUMG00000032594	WFDC2	1
A	A	OTTHUMG00000032614	WFDC3	5
A	A	OTTHUMG00000046411	WFDC5	2
A	A	OTTHUMG00000046354	WFDC6	2
A	A	OTTHUMG00000046342	WFDC8	2
A	A	OTTHUMG00000046332	WFDC9	1
P	A	OTTHUMG00000033071	WISP2	3
P	P	OTTHUMG00000032549	YWHAB	4
A	A	OTTHUMG00000032481	ZHX3	8
P	P	OTTHUMG00000032709	ZNF313	1
A	P	OTTHUMG00000032654	ZNF334	1
A	P	OTTHUMG00000032637	ZNF335	1

P	P	OTTHUMG00000032696	ZNFX1	4
P	A	OTTHUMG00000074023	ZSWIM1	1
P	P	OTTHUMG00000032627	ZSWIM3	1

Table A. 5 Expression profile of 103 protein-coding genes in human chromosome 20q12-13.2 region. The first and sec columns represent the expression status of the genes in HeLa S3 and NTERA-D1 cell lines respectively. A stands for “Absent” which means it is not expressed and “P” stands for “Present” meaning that it is expressed. M stands for “marginal” where no decision can be made for the expression status of the gene

Probe ID	HeLa-S3 STATUS	HeLa S3 Signal	NTERA-D1 STATUS	NTERA-D1 Signal	Matching Transcript 1	Matching Transcript 2	Matching Transcript 3	Matching Transcript 4	Matching Transcript 5	Matching Transcript 6	Matching Transcript 7
236514_at	P	49.3	P	75.8	ACOT8-001	ACOT8-003	ACOT8-004				
204639_at	P	118.6	P	124.3	ADA-001						
218098_at	P	364.8	P	351.6	ARFGEF2-001						
221485_at	P	359.2	P	857.2	B4GALT5-001						
238470_at	P	11.4	P	36.1	C20orf10-001						
209020_at	P	242.4	P	407.4	C20orf111-001	C20orf111-002					
231838_at	P	20.9	P	14	C20orf119-007	C20orf119-009					
228031_at	P	60.3	P	241.7	C20orf121-001						
226805_at	P	157.3	P	69.3	C20orf142-001						
1553960_at	P	20.4	P	103.1	SNX21-011						
218094_s_at	P	93.3	P	170.3	DBNDD2-001	DBNDD2-002	DBNDD2-003	DBNDD2-004	DBNDD2-005	DBNDD2-008	
222044_at	P	137.9	P	77.2	C20orf67-001						
35150_at	P	45.5	P	93	CD40-001	CD40-002	CD40-006				
225031_at	P	41.3	P	134.5	CHD6-001						
201111_at	P	788.9	P	1793.8	CSE1L-001						
221780_s_at	P	203	P	325.6	DDX27-001						
224825_at	P	240.6	P	490.1	DNTTIP1-002	DNTTIP1-004					
55692_at	P	72.4	P	221	ELMO2-001	ELMO2-003	ELMO2-006	ELMO2-007	ELMO2-018		
228307_at	P	140.1	P	119.4	EMILIN3-001						
218709_s_at	P	118	P	403.7	IFT52-001	IFT52-002					
225076_s_at	P	96.5	P	141.2	ZNFX1-001	ZNFX1-004					
213837_at	P	16.6	P	18.8	L3MBTL-003	L3MBTL-005	L3MBTL-010	L3MBTL-011			
201710_at	P	356.7	P	912.2	MYBL2-001						
207700_s_at	P	115.7	P	140.6	NCOA3-001	NCOA3-002					
225145_at	P	123.7	P	323.7	NCOA5-001	NCOA5-002					
217770_at	P	265.7	P	191.2	PIGT-001						
202732_at	P	72.7	P	560.5	PKIG-001	PKIG-002	PKIG-003	PKIG-004	PKIG-006	PKIG-007	PKIG-008
202789_at	P	81.3	P	463.7	PLCG1-005						
200661_at	P	667.7	P	416.6	PPGB-001	PPGB-002	PPGB-004				
209048_s_at	P	172.7	P	794.8	PRKCBP1-001	PRKCBP1-007	PRKCBP1-005	PRKCBP1-006	PRKCBP1-009	PRKCBP1-003	
202071_at	P	402	P	875.7	SDC4-001						

208804_s_at	P	456.8	P	997.9	SFRS6-001	SFRS6-002					
225037_at	P	121.3	P	232.7	SLC35C2-001	SLC35C2-002	SLC35C2-003	SLC35C2-005	SLC35C2-013		
212947_at	P	40	P	100.4	SLC9A8-001						
204433_s_at	P	67.8	P	168.7	SPATA2-001						
211505_s_at	P	273	P	606.2	STAU1-001	STAU1-002	STAU1-003	STAU1-004	STAU1-006	STAU1-007	
211085_s_at	P	31.8	P	52.9	STK4-003	STK4-006					
221471_at	P	564.1	P	624.8	SERINC3-001						
201870_at	P	271.9	P	438.5	TOMM34-001						
208901_s_at	P	825.5	P	1751.5	TOP1-001						
225402_at	P	132.2	P	530.6	TP53RK-001	TP53RK-002					
202954_at	P	1838.5	P	4331.4	UBE2C-001	UBE2C-004					
223186_at	P	268.4	P	279.2	UBE2V1-001	UBE2V1-005					
201002_s_at	P	445.7	P	1772.6	UBE2V1-004						
203892_at	P	14	P	652.8	WFDC2-002						
208743_s_at	P	1550.8	P	1713.6	YWHAB-001	YWHAB-002					
200867_at	P	179.2	P	514.5	ZNF313-001						
228223_at	P	46.5	P	68.9	ZSWIM3-001						
220540_at	P	82.1	A	0.4	KCNK15-001						
41469_at	P	34.6	A	12.9	PI3-001						
203021_at	P	2732.7	A	22.1	SLPI-001						
206318_at	P	12	A	1	SPINLW1-001						
205792_at	P	1709.3	A	14.2	WISP2-001	WISP2-002	WISP2-003				
224909_s_at	M	21.4	P	75.2	PREX1-001						
228737_at	A	2.3	P	26.9	C20orf100-001	C20orf100-002	C20orf100-004				
1569679_at	A	14.4	P	48.1	CDH22-001						
219668_at	A	12.1	P	92.6	GDAP1L1-001	GDAP1L1-004					
218559_s_at	A	1.6	P	543.9	MAFB-001						
203936_s_at	A	15.2	P	69.4	MMP9-001						
202075_s_at	A	6.5	P	293.4	PLTP-001						
210702_s_at	A	1	P	15.8	PTGIS-001						
221377_s_at	A	14.3	P	8.8	RBPSUHL-001	RBPSUHL-002	RBPSUHL-003				
233555_s_at	A	10.5	P	43.9	SULF2-001	SULF2-003	SULF2-005				
206442_at	A	0.8	P	66.1	SEMG1-001						
205244_s_at	A	1.4	P	87.6	SLC13A3-001	SLC13A3-003	SLC13A3-004				
219480_at	A	23.4	P	50.5	SNAI1-001						

220022_at	A	0.6	P	73.8	ZNF334-001						
78330_at	A	8.8	P	15.5	ZNF335-001						
226670_s_at	A	39.8	M	101.2	C20orf119-001						
1553752_at	A	1.2	A	0.7	C20orf165-001						
1570327_at	A	10.3	A	8.1	C20orf62-001	C20orf62-002					
232820_s_at	A	1.9	A	3.6	FAM112A-001	FAM112A-002					
209692_at	A	1	A	22.6	EYA2-003	EYA2-005					
243652_at	A	0.8	A	1.8	EYA2-003	EYA2-004	EYA2-005				
216889_s_at	A	4.7	A	4.4	HNF4A-002	HNF4A-003					
220385_at	A	1.9	A	0.9	JPH2-001	JPH2-002					
211006_s_at	A	9.3	A	18.3	KCNB1-001						
207366_at	A	4.1	A	46.2	KCNS1-001						
232966_at	A	2.3	A	4.1	LPIN3-001	LPIN3-003					
230283_at	A	10	A	40.4	NEURL2-001						
205948_at	A	1.9	A	3.6	PTPRT-001	PTPRT-002	PTPRT-003	PTPRT-004	PTPRT-005	PTPRT-006	PTPRT-007
234774_at	A	9	A	3.9	R3HDML-001						
233299_at	A	17	A	22.8	RIMS4-002						
216030_s_at	A	6.8	A	6.4	SEMG2-001						
220357_s_at	A	15.8	A	16.8	SGK2-001	SGK2-004	SGK2-005	SGK2-006			
210040_at	A	5.1	A	20.8	SLC12A5-002						
215503_at	A	1.1	A	1	SPINT3-001						
205388_at	A	5.7	A	2.3	TNNC2-001	TNNC2-003					
233913_at	A	2.5	A	2.5	WFDC10A-001						
1552999_a_at	A	0.9	A	4.2	WFDC10B-001						
1552608_at	A	1.1	A	9.3	WFDC11-001	WFDC11-002					
1553081_at	A	1.7	A	17.8	WFDC12-001						
1553052_at	A	7.4	A	1.5	WFDC13-001						
232602_at	A	41.7	A	2	WFDC3-001	WFDC3-002	WFDC3-003				
242204_at	A	3.1	A	12.1	WFDC5-001	WFDC5-002					
1552396_at	A	2.1	A	1.4	WFDC6-001	WFDC6-002					
1554156_a_at	A	3	A	0.7	WFDC8-001	WFDC8-002					
1552415_a_at	A	20.4	A	12	WFDC9-001						
217367_s_at	A	18.3	A	37.1	ZHX3-007	ZHX3-008					
223607_x_at*	P	52.2	A	50.5	ZSWIM1-001						
221024_s_at	A	0.3	P	181.6	SLC2A10-001						

207123_s_at	A	10.2	P	31.8	MATN4-001	MATN4-002	MATN4-003				

Table A. 6 Expression Profile of 198 transcripts in human 20q12-13.2 using Affymetrix U133 Plus 2 Expression Arrays on HeLa S3 and NTERA-D1 cell line total RNA. First column denotes the probe name, sec and fourth columns represents whether the gene is expressed or not in HeLa S3 or NTERA-D1 respectively. “A” stands for “absent and “P” stands for present. Third and fifth columns represent the signal intensity of the probe. Transcripts that are represented by each probe are listed in the remaining columns. ”*” The probe for SWIM1 gene is an x_at type probe although it has no cross-hybridization anywhere in the human genome.

HUGO Transcript ID	**Response to Enhancer in HeLa S3**	**Response to Enhancer in NTERA-D1**	**Transcript Type**	**Number of Sp1 Binding Sites**	**Putative Sp1 Binding Site Coordinates Relative to TSS at +1**												
WISP2-001	NR	NR	AT	1	-235												
FAM112A-002	NR	NR	AT	1	-99												
DDX27-005	NR	NR	AT	1	47												
SEMG2-001	NR	NR	RT	1	12R												
WFDC3-006	NR	NR	AT	1	-45												
PPGB-002	NR	NR	RT	2	-149	-67											
SGK2-004	NR	NR	RT	2	-73	76											
SPINLW1-001	NR	NR	RT	2	-130R	3											
UBE2V1-004	NR	NR	AT	3	-148	-113	-73										
PPGB-001	NR	NR	AT	3	-121	-40	-9										
C20orf100-001	NR	NR	AT	4	-118R	-93	-46	-40R									
SGK2-002	NR	NR	AT	4	-224R	-183	-3	35R									
UBE2C-004	NR	NR	AT	5	-182R	-116	-84	-63R	22								
ELMO2-003	NR	NR	RT	6	-226	-182R	-84	-63	-53	35							
R3HDML-001	R	NR	RT	3	-225R	-138	-20										
DNTTIP1-005	R	NR	AT	7	-275R	-239R	-152	-141R	-63	6R	14						
ELMO2-004	NR	R	AT	0													
PRKCBP1-007	NR	R	AT	1	-189												
C20orf100-002	NR	R	AT	1	58												
SEMG1-001	NR	R	RT	2	-157	-21R											
PRKCBP1-001	NR	R	RT	2	-123	-97											
SULF2-001	NR	R	RT	3	-170R	-144R	-79R										
SLC35C2-006	NR	R	AT	3	-183R	-32R	18										
C20orf111-002	NR	R	AT	4	-202R	-141	-89R	-79R									
IFT52-002	NR	R	RT	5	-146R	-123R	-41R	-21R	28								
RBPSUHL-001	NR	R	RT	6	-180R	-133	-114R	-97R	-79	-10							
ZNF313-001	NR	R	RT	6	-107	-100R	-87R	-70R	-56R	47							
ARFGEF2-001	NR	R	RT	7	-240R	-220	-140	-116	-38R	-29	-22R						
KCNB1-001	NR	R	RT	10	-318	-267	-221R	-211	-203R	-187	-162	-139R	-112	-71			
SLPI-001	R	R	RT	0													

ZNF335-001	R	R	RT	0													
MATN4-001	R	R	RT	0													
ZNF334-001	R	R	RT	1	-150												
ACOT8-001	R	R	RT	1	-189R												
ZSWIM3-001	R	R	RT	1	-51												
ZNFX1-003	R	R	AT	1	-213												
CSE1L-001	R	R	RT	2	-76	-51R											
SERINC3-001	R	R	RT	2	-142R	-87											
UBE2V1-003	R	R	AT	2	-108	-5R											
WFDC12-001	R	R	RT	2	-135R	-63R											
CDH22-001	R	R	RT	3	-124R	-50R	43R										
PI3-001	R	R	RT	3	-99R	-87	-33R										
NCOA3-002	R	R	RT	4	-278	-186R	-164	-104									
TOP1-001	R	R	RT	4	-243	-179	-131	6R									
PKIG-001	R	R	RT	4	-133R	10R	18	53									
SLC35C2-005	R	R	RT	4	-143R	-127R	-98R	-38R									
SPATA2-001	R	R	RT	4	-155R	-84R	-67R	64R									
SLC9A8-001	R	R	RT	4	-175R	-140R	-118R	-103R									
UBE2V1-002	R	R	AT	4	-220R	-72	-15	12									
KCNS1-001	R	R	RT	4	-149R	-71R	-39R	-11R									
CD40-002	R	R	RT	5	-111R	-69R	-56	-39	60								
HNF4A-001	R	R	RT	5	-241R	-220	-59R	71R	91								
L3MBTL-003	R	R	AT	6	-272	-222	-174	-124	-73	5							
NCOA5-001	R	R	RT	6	-228	-146R	-102R	-27R	-8R	17R							
SNX21-010	R	R	RT	6	-223	-75	-63	29	49	58R							
SFRS6-001	R	R	RT	7	-224	-81R	-69R	-26	-13R	3	25R						
C20orf111-001	R	R	RT	8	-169	-158	-147	-136R	-88	-66	-20	-13R					
C20orf121-001	R	R	RT	8	-136R	-71	-35R	-26R	3	17	64R	86R					
YWHAB-001	R	R	RT	8	-227R	-207R	-186R	-172R	-161R	-112R	-98	-50					
B4GALT5-001	R	R	RT	8	-80R	-60R	-48R	-29	-14R	37	77	119					
PLTP-002	R	R	AT	8	-208	-188	-168	-114	-71	-37	-19	38					
CHD6-001	R	R	RT	8	-292R	-252R	-228R	-218R	-186	-157	-136	-54R					
C20orf67-001	R	R	RT	8	-243R	-235R	-209R	-194R	-143	-32	12R	57					
GDAP1L1-005	R	R	RT	9	-270	-242R	-215	-197	-185	-147R	-106	-88	-20R				

PLTP-001	R	R	RT	9	-244	-235	-115R	-99R	-86	-63	-52	-17R	22R				
LPIN3-001	R	R	RT	9	-243	-218	-204	-182R	-150	-132	-103	-93	38				
C20orf119-017	R	R	RT	10	-177R	-167	-107	-83	-66R	-37	-21	-13R	9R	26			
ZNFX1-001	R	R	AT	10	-247R	-230R	-183R	-174	-131R	-118R	-88R	-78R	-33	64R			
UBE2V1-001	R	R	RT	10	-230	-206	-192	-176R	-161R	-135R	-68R	-29R	0	11			
SLC12A5-002	R	R	RT	10	-202R	-176R	-140R	-93	-44	33	45	56	68R	82			
PLCG1-005	R	R	RT	13	-244R	-226R	-194	-183	-171	-161R	-116	-95R	-74R	-26R	-14R	48R	69

Table A. 7 Sp1 binding site coordinates of 71 candidate promoter sequences together with their activation status in synergy with SV40 enhancer in HeLa S3 and NTERA-D1 cells. Coordinates are given relative to TSS assigned to +1 and R denotes that the motif is located on the opposite strand of the direction of the transcription.

	HeLa S3		**NTERA-D1**		**Transcription Factor Binding Site Coordinates relative to TSS**						
HUGO Transcript ID	**Promoter Activity**	**Response to Enhancer**	**Promoter Activity**	**Response to Enhancer**	**Transcript Type**	**AP1**	**Sp1**	**NFKB1**	**CREB**	**YY1**	**MYC**
SNX21-010	-6.603	9.658	3.573	29.175	RT		-223;-75;-63;49;58R;	-160;-160R; -105;	-233R;		
C20orf67-001	-15.971	0.786	-21.828	1.533	RT		-235R;-209R;-143;-32;			24;	
CDH22-001	-13.116	21.891	0.060	31.387	RT		-124R;-50R;43R;	15R;	-226R; -134R;		
CHD6-001	-8.434	8.059	0.633	4.102	RT		-292R;-252R;-228R;-218R; -186;-157;-136;			-33;	-338;-215R; -214;
DNTTIP1-005	-5.036	2.977	-50.950	-10.642	AT	-230	-152;6R;14;	-42R;	-257R;	-250R;23;	-258;
HNF4A-001	-12.118	52.349	-44.500	17.122	RT		-220;91;				
KCNS1-001	-45.841	12.657	0.145	16.039	RT			-158R;			
ZNFX1-003	-51.849	12.698	-66.822	17.642	AT			-208;		1;	
LPIN3-001	-33.133	65.502	0.627	40.684	RT	51	-243;-218;-204;-182R;-150; -132;	-125;			
MATN4-001	-43.471	15.320	0.037	32.157	RT					-91R;	
PI3-001	-60.178	21.980	-99.632	1.246	RT		-99R;-87;-33R;	-114R;			
PLCG1-005	-64.626	3.486	0.061	15.622	RT		-171;-161R;-116;-95R;-74R; -14R;48R;				-71;
PLTP-002	-24.783	9.568	1.754	11.203	AT	67	-208;-188;-168;-114;-37;				
R3HDML-001	-53.963	0.363	-71.446	-36.871	RT		-225R;-138;-20;			-100R;	
SLC12A5-002	-16.852	2.160	0.280	5.753	RT		-202R;-93;-44;33;45;			-241R;	
UBE2V1-001	-37.261	9.265	0.758	34.093	RT		-230;-206;-192;-176R;-29R;		-51R;		-187R;
UBE2V1-003	-10.475	9.693	12.976	48.084	AT		-108;-5R;	-12;	-122R;		
WFDC12-001	-59.061	2.959	-59.858	0.293	RT			-149R;		-224R; -121R;13;	-232R;-231;
ZNF335-001	-15.852	0.642	2.499	5.279	RT	14			-250R; -248;		
ZSWIM3-001	-3.640	0.937	2.195	30.418	RT		-51;			-28;	

Table A. 8 Transcription Binding Site Profile of candidate promoters that showed activity only in synergy with SV40 enhancer in HeLa S3 cells. Binding site coordinates are given in relative to TSS at +1. Transcripts coloured yellow denotes the ones that gave activity only in synergy to enhancer in both cell lines. All promoter activities are scaled from 0 to 100.

	HeLa S3		**NTERA-D1**			**Transcription Factor Binding Site Coordinates relative to TSS**					
HUGO Transcript ID	**Promoter Activity**	**Response to Enhancer**	**Promoter Activity**	**Response to Enhancer**	**Transcript Type**	**AP1**	**Sp1**	**NFKB1**	**CREB**	**YY1**	**c-myc**
C20orf100-002	-45.973	-19.161	-22.819	1.521	AT			-133;-7;-7R;	-219;		16;
C20orf111-002	-44.335	-17.244	-7.465	0.245	AT		-141;-79R;	-180;-56;			
C20orf67-001	-15.971	0.786	-21.828	1.533	RT		-235R;-209R; -143;-32;			24;	
ELMO2-004	-18.330	-2.533	-51.339	2.406	AT						38R;
HNF4A-001	-12.118	52.349	-44.500	17.122	RT		-220;91;				
ZNFX1-003	-51.849	12.698	-66.822	17.642	AT			-208;		1;	
PI3-001	-60.178	21.980	-99.632	1.246	RT		-99R;-87;-33R;	-114R;			
PRKCBP1-007	-58.050	-38.907	-76.783	2.558	AT		-189;				
RBPSUHL-001	-60.823	-27.486	-43.882	1.053	RT		-133; -97R;-79;	-126R;			
SULF2-001	-21.963	-3.153	-14.019	3.982	RT	-89	-79R;		-163;	-31R;	
SLC35C2-006	-47.335	-41.458	-43.198	0.584	AT	-21	-183R; -32R;18;				
WFDC12-001	-59.061	2.959	-59.858	0.293	RT			-149R;		-224R; -121R;13;	-232R;-231;

Table A. 9 Transcription Binding Site Profile of candidate promoters that showed activity only in synergy with SV40 enhancer in NTERA-D1 cells. Binding site coordinates are given in relative to TSS at +1. Transcripts coloured yellow denotes the ones that gave activity only in synergy to enhancer in both cell lines. All promoter activities are scaled from 0 to 100.

	HeLa S3		NTERA-D1			Transcription Factor Binding Site Coordinates relative to TSS					
HUGO Transcript ID	**Promoter Activity**	**Response to Enhancer**	**Promoter Activity**	**Response to Enhancer**	**Transcript Type**	**AP1**	**Sp1**	**NFKB1**	**CREB**	**YY1**	**MYC**
C20orf100-001	-50.180	-60.975	-65.563	-92.177	AT		-118R;-93;		-70R;		
FAM112A-002	-26.372	-93.324	-35.584	-73.521	AT						
DDX27-005	-24.872	-80.994	-29.494	-45.016	AT						
PPGB-001	-9.867	-36.907	-14.991	-0.713	AT		-121;-40;	33;		4;	-162R;
PPGB-002	-64.579	-100.0	-66.191	-90.674	RT		-67;				
SEMG2-001	-99.896	-69.880	-52.575	-65.664	RT						
SGK2-002	-19.588	-77.879	-27.874	-53.351	AT		-224R;	-86R;-46;-46R;		-9;43R;	
SGK2-004	-30.894	-58.041	-55.173	-20.228	RT			-103R;			
SPINLW1-001	-23.233	-34.263	-37.073	-42.911	RT		-130R;			19R;	
UBE2C-004	-51.751	-20.342	-61.014	-14.463	AT		-84;-63R;22;	-109;-69;-69R;		-35;	
UBE2V1-004	-10.536	-38.391	-23.355	-42.128	AT		-148;				-197R;
WISP2-001	-49.812	-64.179	-65.818	-80.927	AT		-235;			-81;	

Table A. 10 Transcription Binding Site Profile of candidate promoters that did not show in any cell line with or without enhancer. Binding site coordinates are given in relative to TSS at +1. All promoter activities are scaled from 0 to 100.

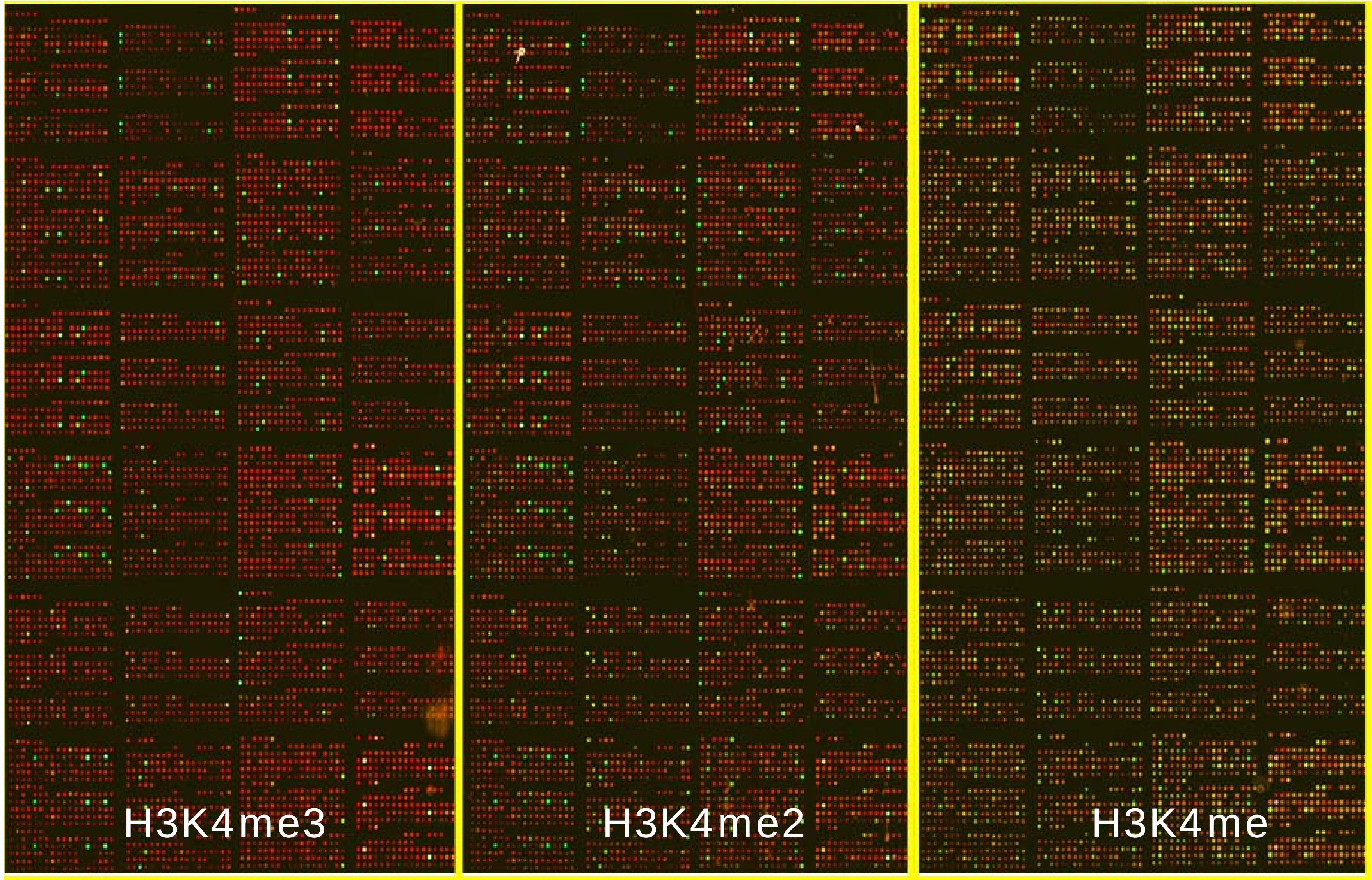

Figure A1. The scanned arrays of H3K4me3, H3K4me2 and H3K4me hybridizations. Spots that are enriched with the antibody bound DNA fragments are green.

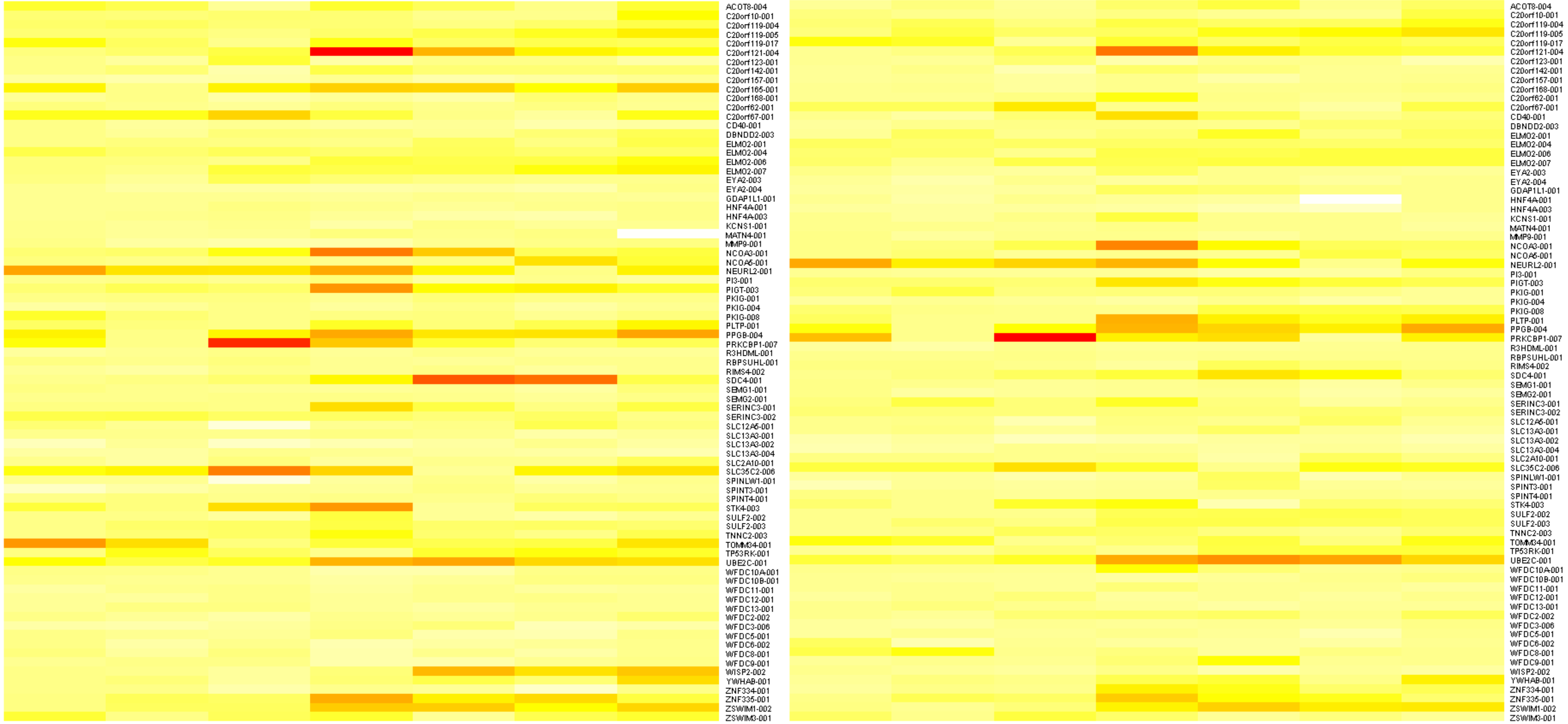

Figure A2. The heat maps that displays the spot intensities of all 79 transcripts represented on the array (see section 5.3), corresponding to 66 genes in HeLa S3 cell (left) and NTERA-D1 cells (right)

Cell Type	Start	End	Length	Spot St.	Spot End	Spot Information	polII	H3K4me	H3K4me2	H3K4me3	H3ac	H4Ac	CTCF
Jurkat	42671737	42672051	314	42670305	42672476		0.64	1.74	1.84	0.31	1.30	2.30	6.59
NCCIT	42671747	42672056	309										
FHs738Lu	43131582	43131910	328	43130169	43132851		0.42	1.79	0.26	0.04	0.11	0.83	0.08
FHs738Lu	43132024	43132330	306										
FHs738Lu	43158352	43159058	706	43158018	43159290		-0.12	1.00	0.45	-0.02	1.26	0.61	3.20
HepG2_PolyA-	43360758	43361072	314	43360657	43363032		-1.00	-0.56	-1.58	-0.32	-2.08	-0.43	1.73
HepG2_PolyA+	43401492	43402875	1383	43400608	43402868		2.08	4.18	0.86	0.10	0.92	0.62	-0.04
HepG2_PolyA-	43401512	43402875	1363	43402560	43405196		3.01	5.42	7.82	1.61	4.42	1.30	-0.06
HepG2_PolyA+	43403188	43404327	1139	43403941	43405207		0.50	1.04	1.79	0.47	0.61	0.42	0.16
HepG2_PolyA-	43403188	43404327	1139										
HepG2_PolyA+	43404378	43404818	440	43403941	43405207		0.50	1.04	1.79	0.47	0.61	0.42	0.16
HepG2_PolyA-	43404378	43404795	417										
NCCIT	43902419	43902983	564	43901741	43903780	3' end of ACOT8 gene	1.28*	0.02	-0.51	-0.17	0.04	0.00	0.01
A375	43902484	43902983	499										
Jurkat	43902484	43902983	499										
SK-N_AS	43902484	43902983	499										
HepG2_PolyA+	43902489	43902983	494										
PC3	43902489	43902983	494										
HepG2_PolyA+	44047015	44047509	494	44046483	44048188		-0.23	0.04	-0.16	-0.05	0.55	0.19	1.74*
A375	44047020	44047509	489										
NCCIT	44047020	44047509	489										
Jurkat	44047025	44047509	484										
SK-N_AS	44047025	44047509	484										
FHs738Lu	44047040	44047509	469										
U87	44047040	44047509	469										
HepG2_PolyA-	44047045	44047509	464										
PC3	44047045	44047509	464										
A375	44047581	44048239	658										
HepG2_PolyA+	44047581	44048239	658										
HepG2_PolyA-	44047581	44048239	658										
Jurkat	44047581	44048239	658										
NCCIT	44047581	44048239	658										
PC3	44047581	44048239	658										

SK-N_AS	44047581	44048239	658										
U87	44047581	44048239	658										
U87	44147844	44148155	311	44147270	44149246		1.83	2.73	2.91	0.30	0.93	0.20	0.93
PC3	44147850	44148180	330										
HepG2_PolyA-	44270148	44270483	335	44269828	44270823		2.11	3.67	8.15	2.50	4.79	3.57	-0.09
A375	44298702	44299421	719	44298319	44300488		-0.10	-0.00	-0.55	-0.01	-0.27	-0.03	-0.07
FHs738Lu	44298702	44299426	724										
HepG2_PolyA+	44298702	44299416	714										
HepG2_PolyA-	44298702	44299426	724										
PC3	44298702	44299421	719										
U87	44298702	44299426	724										
HepG2_PolyA+	44382036	44382579	543	44380475	44382544	5' end of GENSCAN37458	0.73	2.77	3.83	1.25	1.02	1.00	-0.05
Jurkat	44382036	44382579	543										
PC3	44382036	44382586	550										
A375	44382147	44382569	422										
U87	44382147	44382558	411										
A375	44820422	44820843	421	44820067	44822119		-0.20	-0.17	0.08	-0.03	0.08	-0.19	6.63
U87	44820422	44820843	421										
SK-N_AS	45034269	45034740	471	45034995	45037048		-0.13	-0.08	0.22	-0.06	-0.50	-0.18	3.28
HepG2_PolyA+	45035321	45035682	361	45034995	45037048		-0.13	-0.08	0.22	-0.06	-0.50	-0.18	3.28
NCCIT	45035321	45035677	356										
PC3	45035326	45035677	351										
A375	45035336	45035667	331										
Jurkat	45035336	45035682	346										
SK-N_AS	45035336	45035677	341										
FHs738Lu	45035351	45035677	326										
U87	45035356	45035662	306										

Table A 11a. The coordinates and cell specificity of the transfrags obtained in the study by Cheng J et al, that showed enrichments with antibodies used in this study in HeLa S3 cells. * denotes a signal which is reported here although it is below the threshold used in this study.

Cell Type	Start	End	Length	Spot St	Spot End	Spot Information	polII	H3K4me	H3K4me2	H3K4me3	H3ac	H4Ac	CTCF
Jurkat	42671737	42672051	314	42670305	42672476		0.18	0.69	1.58	0.26	0.68	0.17	4.10
NCCIT	42671747	42672056	309										
FHs738Lu	43158352	43159058	706	43158018	43159290		0.01	0.06	0.10	-0.11	0.52	0.02	3.20
HepG2_PolyA+	43401492	43402875	1383	43400608	43402868		0.66	2.08	2.03	0.23	1.18	0.60	0.03
HepG2_PolyA-	43401512	43402875	1363	43402560	43405196		2.24	1.00	13.85	12.06	2.65	0.13	0.05
HepG2_PolyA+	43403188	43404327	1139	43403941	43405207		0.27	2.03	3.18	3.15	1.22	0.06	0.59
HepG2_PolyA-	43403188	43404327	1139										
HepG2_PolyA+	43404378	43404818	440	43403941	43405207		0.27	2.03	3.18	3.15	1.22	0.06	0.59
HepG2_PolyA-	43404378	43404795	417										
NCCIT	43902419	43902983	564	43901741	43903780	3' end of ACOT8 gene	1.90	-0.16	-0.24	-0.32	0.41	0.21	0.10
A375	43902484	43902983	499										
Jurkat	43902484	43902983	499										
SK-N_AS	43902484	43902983	499										
HepG2_PolyA+	43902489	43902983	494										
PC3	43902489	43902983	494										
A375	43903806	43904471	665	43903837	43906350	3' end of C20orf161-001 gene	2.05	-0.26	-0.08	-0.05	0.33	0.27	0.13
FHs738Lu	43903806	43904116	310										
HepG2_PolyA+	43903806	43904471	665										
Jurkat	43903806	43904471	665										
NCCIT	43903806	43904471	665										
PC3	43903806	43904471	665										
SK-N_AS	43903806	43904471	665										
A375	44298702	44299421	719	44298319	44300488		-0.01	0.44	2.67	0.35	0.67	0.04	-0.04
FHs738Lu	44298702	44299426	724										
HepG2_PolyA+	44298702	44299416	714										
HepG2_PolyA-	44298702	44299426	724										
PC3	44298702	44299421	719										
U87	44298702	44299426	724										
NCCIT	44737903	44738536	633	44736072	44738637		0.61	1.22	2.02	0.20	0.84	0.33	-0.03
PC3	44764446	44764757	311	44764688	44766247		0.09	0.83	1.88	0.16	0.51	0.03	-0.01
A375	44820422	44820843	421	44820067	44822119		-0.16	0.26	0.34	0.03	-0.07	-0.05	5.39
U87	44820422	44820843	421										
SK-N_AS	45034269	45034740	471	45034995	45037048		-0.04	-0.06	-0.08	-0.09	-0.22	0.12	3.50
HepG2_PolyA+	45035321	45035682	361	45034995	45037048		-0.04	-0.06	-0.08	-0.09	-0.22	0.12	3.50

NCCIT	45035321	45035677	356										
PC3	45035326	45035677	351										
A375	45035336	45035667	331										
Jurkat	45035336	45035682	346										
SK-N_AS	45035336	45035677	341										
FHs738Lu	45035351	45035677	326										
U87	45035356	45035662	306										
A375	45321779	45322470	691	45319908	45322097		1.64	-0.04	-0.52	-0.53	-0.03	0.37	-0.12
PC3	45321779	45322470	691										
SK-N_AS	45321779	45322465	686										
HepG2_PolyA+	45321784	45322465	681										
U87	45321799	45322465	666										
NCCIT	45321804	45322455	651										
FHs738Lu	45321824	45322285	461										
Jurkat	45321824	45322465	641										
HepG2_PolyA+	45343564	45343975	411	45341963	45344242		1.78	0.52	-0.42	-0.48	0.18	0.21	-0.08
HepG2_PolyA+	45346308	45346638	330	45345737	45348010		1.63	0.01	-0.03	-0.02	0.21	0.32	-0.07

Table A 11b. The coordinates and cell specificity of the transfrags obtained in the study by Cheng J et al, that showed enrichments with antibodies used in this study in NTERA-D1 cells.

APPENDIX B

B1. Primers for Amplification of Putative Promoter Fragments (See Chapter 4)

Primer Name (20ace database)	Primer1	Primer2
ADA_80509	AAGGATCCCCGGCCACGCTCTTCTTAAC	AAGCTAGCGCCGGGGGGAAGGCAGCGCC
C20ORF100_79329	AAGCTAGCCATGCACTGAGTACAGGTAT	AAGGATCCCCCCATCCCATTCGTATAAG
C20ORF111_79334	AAGGATCCGCTGCAGCCAAAGCACAAAG	AAGCTAGCTCGGCCCCACTCCTTTCTTC
C20ORF121_79370	AAGCTAGCTTCCTTCCTTCTCTGAGCCT	AAGGATCCTCCCTCTCCTCAGCTCCCAG
C20ORF169_79455	AAGCTAGCTCCTGTCTGTCAGCGCTGTT	AAGGATCCGTCTGCATGAGGACGATCTG
C20ORF35_79441	AAGCTAGCCCGGCGTCGGGGGGCGCGGC	AAGGATCCTTGATGCTCTGCGCCTTTTG
ELMO2_80468	AAGGATCCTCTCTTCCCTGCGTCTTTCG	AAGCTAGCGTTTCTTTAAGAGGCCTCCA
KCNS1_80507	AAGGATCCCTCTCCTTCCTTCTTCCTTG	AAGCTAGCGTACCCTGGCACATACTAAA
KIAA1404_79649	AAGGATCCATGCCCAGTACATGGAGAAC	AAGCTAGCAACCGCAGACCACCATGGTC
LPIN3_80393	AAGCTAGCGGAGGTGGGGCAGGCGGTGG	AAGGATCCAGAGGGTCATCTCTTCTCCC
MATN4_80334	AAGGATCCAGCAAAGAAGGCCTCTCATG	AAGCTAGCCAGAGCAAGACTCTATCTAA
OVCOV1_80365	AAGGATCCGAATCTCCCCACCGGAAGTG	AAGCTAGCCAGACAATTACACGGCTGTG
OVCOV1_80366	AAGGATCCCACCGCCTTCCACAAAAAGG	AAGCTAGCGGGACCTGTGAGCGCTGCAT
PRKCBP1_79596	AAGGATCCATGGAATGCTGTCATGAGCC	AAGCTAGCATGTTTCTGAAAGTGTTTTC
TOMM34_79390	AAGGATCCCGAGTTGGGAGCTCCTTCCT	AAGCTAGCAAATGAGAGTCTTCGTTTGC
UBE2C_80312	AAGCTAGCTGATGCGGCCTACCACTCGC	AAGGATCCTTGAGCGTCTTCATCTAGGC
UBE2V1_80529	AAGGATCCAGATCACGCCGCCCAATGAC	AAGCTAGCCGACGCGCCGCGAGCTCCCA
UBE2V1_80531	AAGGATCCTACTCGCATGGATTCCTTGG	AAGCTAGCAACCATTTTAATCTTCCCCA
ZNF335_79553	AAGGATCCTAACTCTACCCGAAGCTCAC	AAGCTAGCATGACGTCACAACCACTTCC
ARFGEF2_79627_S	AAGAGCTCGCCTTAGGCTCCGCCCCGCT	AAGGATCCTCGGCTAGGATCTTCTCCAG
B4GALT5_80543_S	AAGGATCCTAGAGAGCCAGGCCGGGCCT	AAGAGCTCGATTACTCCATGCGGGCTGC
BA269H4.1_79623_S	AAGGATCCCTGGGCGCCTCCATTCTAG	AAGAGCTCGTGGCACTTGGCAGACGGTC
C20ORF100_79330_S	AAGAGCTCGGCTTGCGTGCCTGGAACCC	AAGGATCCACGAAAAGCCTGCGAGGTTG
C20ORF122_79408_S	AAGGATCCTCACAAGAACGAGAGACACC	AAGAGCTCCTGGTGGCGCCTAATGGTCA
C20ORF130_80415_S	AAGGATCCTGCCGGTATTCAGCCTTTAC	AAGAGCTCCACGGAGGCTGGAATCAGAG
C20ORF161_79534_S	AAGAGCTCTGGGACAGATTGGAGATGAC	AAGGATCCTGCGTCCCACGGTGCATTCA
C20ORF164_79540_S	AAGAGCTCCATACACAGAACCTGACTCT	AAGGATCCTGGCCTATTCTGGAACTCAG

C20ORF169_79457_S	AAGAGCTCGAACTCCTGACCTTGTGATC	AAGGATCCTATTCCTACTCCAGCAGACG
C20ORF170_79482_S	AAGGATCCTTCCCTATGGAGACAGCTTC	AAGAGCTCCTCCGTGTGGGCATCAGAGC
C20ORF1901_01027_S	AAGGATCCCCATTCTTGACGCGGAGTAC	AAGAGCTCCCAAGCCCGCCTCGCGCTGC
C20ORF35_79439_S	AAGAGCTCGCATCCGAGCGAGCGGAGAC	AAGGATCCTCTGACTCAGTCCGTCCGCT
CDH22_80491_S	AAGGATCCTTCCTGCGCAGAAAGGATGC	AAGAGCTCTAGGAGATTGGGAGTACAGG
CHD6_79270_S	AAGGATCCCAAGGACACCCTAACCCATC	AAGAGCTCCCGCCCCCTCGGCCCGCGGG
ELMO2_80469_S	AAGGATCCTGGTCGATTTCAAGGAGCTG	AAGAGCTCTGTCCCATTTCTTCTTGCTG
EYA2_80324_S	AAGAGCTCGTGAGCGCGCTCTGCTCGGC	AAGGATCCTACTCACCGCGCGCAGTGGAG
EYA2_80326_S	AAGAGCTCTGTTCTAGCTGAGTCTCTCA	AAGGATCCTTCCTGAGCCTGCAGCCGAG
HNF4A_79362_S	AAGAGCTCCTTCTGCTCCGGCCCTGTCC	AAGGATCCTTCACGCTGACCATGGCCAA
KCNB1_80374_S	AAGGATCCTGCGAGGCAGAAAGCGAAAG	AAGAGCTCGGCTCCAGGGGCATCTCTTG
KCNK15_79378_S	AAGAGCTCGGGAGGAAGGGAGGGGAGGG	AAGGATCCTCCCAACCTGCTCCGTGTG
KIAA1404_79647_S	AAGGATCCGCGAGGCAAGGAAGAATCAG	AAGAGCTCGGTCCGCTCCCCTCAGCCGC
L3MBTL_79296_S	AAGAGCTCCTCTGGGACTGGGCCTGTGG	AAGGATCCTACTCGCTTATCCGGAACTG
C20ORF142_79342_S2	AAGGATCCATTATTCCTAAGGAAACTTC	AAGCTAGCCTCCTCTCCCCGAGCCTC
R3HDML_79344_S	AAGAGCTCAATATGACTTGGGGTAAGGT	AAGGATCCCTGAGCGTTCTCCAACAGTG
TDE1_80544_S	AAGGATCCTCCTGAGGCTGCTTTCTAAC	AAGAGCTCGCCTCTCTGTACACTTCTGT
TOP1_80397_S	AAGAGCTCCTGCTTGGGGTGGGACGCCG	AAGGATCCAGACTCCAGAAACGGCTGAG
YWHAB_79386_S	AAGAGCTCCCCACCTCCTTTTCCCGCTC	AAGGATCCTGACCTCTTCCTTCCTAAGC
WISP2_80484_S2	AAGAGCTCTTCGGTGTGCCTCTCATGTC	AAGGATCCAAGCTGTCACAGGCTCTTGG
CSE1L_80345_S	AAGAGCTCTGCCAAGCCCTGACATCAAA	AAGGATCCTTCAAACCCCGGCAAAATGG
ADA_80509_S2	AAGGATCCAAAGTTGCTGGAGGAGCCGG	AAGAGCTCTCGGTGGCCGCTCGGCTTTC
B4GALT5_80543_S2	AAGGATCCGATTACTCCATGCGGGCTGC	AAGAGCTCTCCCGCAGCTCCCCGTCC
BA269H4.1_79623_S2	AAGGATCCTGGGCGCCTCCATTCTAGC	AAGAGCTCTGCACGCGCAGAGCGCCTCT
C20ORF10_79460_S2	AAGGATCCTGAGCATGGTTTGGGAATCC	AAGAGCTCACCTGCAGCTGTGACATCAC
C20ORF111_79333_S	AAAAGCTTGAGCGTCTCTCCCAAATCTC	AAGAGCTCGGTCGGGAGCAAATCCACGC
C20ORF119_79391_S2	AAGCTAGCTCCCAGTCCTCGGCTCTTTC	AAAAGCTTAAGCAGGAGCCGGGTCACC
C20ORF138_79493_S2	AAGAGCTCAAAAAATTCAATCCTCTGTG	AAGGATCCTTTGACCAGGCACTGCTGTG
C20ORF146_79490_S2	AAGAGCTCATGACTCTGATCAGATCGTG	AAGGATCCCAGTTAATTCTAAGCAACTC
C20ORF162_79537_S2	AAGCTAGCAGTAACGTGTCCTTCCAGTC	AAGGATCCACTTTCACCCCTTCAGGAAG
C20ORF163_79539_S	AAAAGCTTGACCTTGACCCGGCATTGAC	AAGAGCTCCTCCTCAGCGCTTCCGTGCA
C20ORF165_79541_S2	AAGGATCCTTCCATCTCTGGTCATCTGC	AAGAGCTCCGGTTTGTGAGGGAATAGTG
C20ORF167_79502_S2	AAGAGCTCACAACTCCGCTGGACTCTGTC	AAGGATCCCCCCAGCCTACAGAAAACTG
C20ORF167_79503_S2	AAGAGCTCAGATCCGTCACCATCTCTCC	AAAAGCTTTTGGCGTTTTCGCAGCCCGA
C20ORF169_79455_S2	AAGCTAGCGTCTGTCAGCGCTGTTTTGG	AAGGATCCTAGCTGCGGAACTGACCCG

C20ORF169_79457_S	AAGAGCTCGAACTCCTGACCTTGTGATC	AAGGATCCTATTCCTACTCCAGCAGACG
C20ORF170_79482_S	AAGGATCCTTCCCTATGGAGACAGCTTC	AAGAGCTCCTCCGTGTGGGCATCAGAGC
C20ORF190_101027_S	AAGGATCCCCATTCTTGACGCGGAGTAC	AAGAGCTCCCAAGCCCGCCTCGCGCTGC
C20ORF35_79439_S2	AAGAGCTCCATCAAGTGTGCGTGGGCAG	AAGGATCCTGGGGCTGCGCCTCCAGGG
C20ORF64_79581_S2	AAGGATCCGCCAGCTCTGAGTCTCAATT	AAGAGCTCAAATCTTCGGTGACTCTCGC
C20ORF65_79314_S2	AAGGATCCAATCTGCTGAGTGGGATTCG	AAGAGCTCTGCTATAGTTCTGTCTCGGG
C20ORF67_79550_S2	AAGAGCTCTGGTACCGCCTCTGAGGGAC	AAGGATCCTTCTAAGACAGCCCCCACGC
C20ORF9_79318_S2	AAGAGCTCTTTACCTGAGGCCCAGCGAC	AAGGATCCTTTGGGCACTTGGGTCCTAC
CHD6_79270_S2	AAGGATCCTAGCCAGCTACCTGAGAGAC	AAGAGCTCTTCGCCGGTCCAAAACACAG
DDX27_80489_S2	AAGAGCTCGCAGAAATGTGTGCTAAAGC	AAGGATCCAGCTGCATACTGTCTTGCAG
DJ1049616.1_79606_S2	AAGGATCCAATCGCAGGGCAAACCGAG	AAGAGCTCTCCCCTTAGCTAGCAACTCG
dJ688G8.2_79485_S2	AAGGATCCTATCTCGGGTTCTAACTGCC	AAGCTAGCACCTCTTTTGCTGCCACTAC
dJ688G8.5_79486_S2	AAGGATCCCAGCCAAATTCTACCAGAC	AAGCTAGCTCATTTTTTTGCGGGGCAGG
EYA2_80324_S2	AAGAGCTCTTGCAGGGAGGATGTGCTGC	AAGGATCCGTTGCTGTCTCTGCCGTTGC
GDAP1L1_79360_S2	AAGCTAGCATGGGCCACCAGCTGCAGTT	AAAAGCTTTGGCCCAATACACTGAGCTG
HNF4A_79363_S2	AAGAGCTCACCTGCCCTACCCTGGGCG	AAGGATCCCGTGAGTCATGATGCCTGCC
JPH2_80307_S2	AAGGATCCCTTCTTGCACCAAGTTCTCC	AAGCTAGCAGAACACCCAGTCCTGGAAC
KCNB1_80374_S2	AAGGATCCCGTCTTCTCACCTCCATCCC	AAGAGCTCGAGTTTCAGCACTCTAAGGG
L3MBTL_79296_S2	AAGAGCTCTTCGCCTCATGCCAGCTCAC	AAGGATCCAGTTGAAGCACTCCTAGGCC
MMP9_80337_S2	AAGAGCTCTGCCAGAGGCTCATGGTGAG	AAGGATCCCCCACAAGCTCTGCAGTTTG
MYBL2_80408_S2	AAGAGCTCGTCAGGACCCGGGCTGCTC	AAGGATCCCACAACTCCGAAGTAGCGGC
NCOA3_80404_S2	AAGCTAGCTTAAATCGGAAACTCGCCGC	AAAAGCTTAAATTAAGGGCAGGGCTAGG
NCOA5_79559_S	AAGGATCCTAGGACAAAGGCGCCACCAAC	AAGAGCTCTGAAAACAGAAAAAAAGTAC
OVCOV1_80363_S	AAGGATCCATTCTTACCTTACACAATGG	AAGAGCTCGGCATCTTAAGTTTTAGATA
OVCOV1_80365_S2	AAGGATCCGCACACCCACTCACCTCGG	AAGCTAGCGGTTCCAAAGCAAGCCCCTG
PI3_79418_S	AAGAGCTCTCTACTCTGTGAGAAAGTAG	AAGGATCCACACCACCACGATCAAGAAG
PIGT_79434_S	AAGAGCTCCATGAACTAGGTGCGGCCTC	AAGGATCCATGACAAGTTCCTCCCGCAG
PKIG_80411_S2	AAGAGCTCACTGTCCCCCTTTCCGTATC	AAAAGCTTTCCTGCTCACCTGCGGTCTC
PKIG_80412_S	AAGAGCTCAAGGAAGGATATTAGGCAAG	AAGGATCCTCCTCAAAATCCAGTGGTCC
PLCG1_80514_S	AAGAGCTCGCGCTCCCGCCGCCATCGCG	AAGGATCCAGCCGTTGGCGCAAGGGGAC
PLTP_80348_S2	AAAAGCTTTCACGTGGGATGGCGGGCA	AAGAGCTCGTCCCAGCAAAGTGGGATTG
PLTP_80349_S	AAGGATCCTCCTTGACTCCACCTTTCTG	AAGAGCTCGGCAAAGAAGGCCACTTCTA
PPGB_80497_S2	AAGAGCTCAGCTCTTTCTCCTCGATCTC	AAGGATCCGGGAAATGCCTTACAAGGTC
PPGB_80498_S	AAGAGCTCGAGCCAGGAGGGGTCGCTGC	AAGGATCCTCCTCGTGGAACCATATCTG
PPGB_80500_S	AAGAGCTCACTCCCTTCCCCGAGCCTCT	AAGGATCCTGCCAGCTCACCTCTGCTC

PRG5_79407_S	AAGGATCCATTAACACCTGCAGCCTCAG	AAGAGCTCCCTCAACAACCTTCACCCTG
PRKCBP1_79590_S	AAGGATCCAAGTCGAGCTTACCTCTGTG	AAGAGCTCGGGCCCCTTTCCCAAGTTTT
PTE1_80338_S	AAGGATCCTCTAGTTCAATGCTGCAGGC	AAGAGCTCGGTGTATTTGCTATCTGACA
PTGIS_80496_S	AAGGATCCAGCAACAGTGCGGCCAGGAG	AAGAGCTCATCCCTTCCGCCCCCTCCCC
PTPRT_80315_S	AAGGATCCAGCTGCAGCCTCAGGAGCAG	AAGAGCTCAACGGCGGCGGCGTTAGGAC
RBPSUHL_80389_S	AAGAGCTCGCGTGGTGGCGTGGCAGCGA	AAGGATCCTCTGGACGAGTCCAGTGCTG
SDC4_80515_S2	AAGGATCCCTGCCTGGCAGTGGGTCAAA	AAGAGCTCAGCAGCGCGAACAGACGGG
SEMG1_79416_S	AAGAGCTCTGTAAAATTAAAGTGATCTG	AAGGATCCCTCCACTCACCTTTTTGTCC
SEMG2_79417_S	AAGAGCTCAATTACTTTTGTAACCTGAA	AAGGATCCATGTTCCTTGAGTGGGTGTG
SFRS6_79292_S2	AAGAGCTCTCTGCGGCTGGATTAGAGAC	AAGGATCCAAGCCAGTAACCGCGGTGC
SGK2_80383_S	AAGAGCTCGCCAGTCTTGGGTCTCTCTG	AAGGATCCAGCTCCAGAAGTCCAAGATG
SGK2_80385_S	AAGAGCTCAGGCTCAGGGAGCTGAAGTC	AAGGATCCCACCTGAAAACTCCAGGTTC
SLC12A5_79558_S	AAGAGCTCACGCAATCCCCCAGTTTTTG	AAGGATCCCAGGTTGTTTAGCATGGTGG
SLC13A3_80329_S2	AAGGATCCAGTACGCGCCTTAATCCTCG	AAGAGCTCACACCTTCTTGGCCGCTGCT
SLC2A10_79578_S	AAGAGCTCTGGGCGAGGGAGGGGGTCCT	AAGGATCCTGTCTACACCCTGGGAAGGG
SLC9A81_06483_S	AAGAGCTCCCCTGTGATGGGGAAAAGCC	AAGGATCCAAAAAGCCCACTCACTCCTC
SLPI_80494_S2	AAGGATCCGACTTCATGGTGAAGGCAGG	AAGAGCTCTAATGGCCTGGGATCTTGTG
SNAI1_80350_S	AAGAGCTCGCGCTGCGCCAGCGAACCCC	AAGGATCCGTAGTTAGGCTTCCGATTGG
SPATA2_79658_S	AAAAGCTTCGAGGGAAGCAAGCGAGAG	AAGAGCTCGCCTGTTTACTTCCGGGTCC
SPINLW1_79467_S2	AAGGATCCGAGGCTCAAAAGTCCAGAAG	AAGAGCTCTCTGAAGGTAGCCTGGAAAG
SPINLW1_79468_S	AAGGATCCTAGTATCACTTACCCCGGAG	AAGAGCTCAGTTACTTAAATTTCATGGG
STAU_79633_S	AAGGATCCAAACGCTGAAGAGCCGCTCA	AAGAGCTCAGTCCAGCAGGCAGCGCGAA
TIX-1_79257_S2	AAGGATCCTCTGAAGGCTGGAATTCTGG	AAGAGCTCAAGGCAGAAGCTCATCTTTC
TIX-1_79258_S	AAGGATCCCAGCTGCACCAACCGACTC	AAGAGCTCACTCGAGGAGGCTGACTGATG
TIX-1_79259_S	AAGGATCCATGCTGGCATCTTGCAACAC	AAGAGCTCTCATATGCGGGAAGGCTATG
TNFRSF5_80377_S	AAGAGCTCCCTGGGGGCAAAGAAGAAGA	AAGGATCCGGGCAAAAACAACTCACAGC
TNNC2_79522_S	AAAAGCTTACTGGGATGACTCTGCGGCA	AAGCTAGCCGGAGGGGCTCAGGACCCTC
TNNC2_79524_S	AAGGATCCAAGTCCCCTCTTGTCCTTAC	AAGAGCTCACTCCAAGGCAGTGGAACAG
UBE2V1_80530_S	AAGGATCCTCTTGCGTCGCTCTTGCTTG	AAGAGCTCTCACAATCGCGCCTTCCCAA
UBE2V1_80532_S2	AAGGATCCATGTTTAGCAGCGGCAGCAG	AAGAGCTCGGGTCATTTCGGCTCATCTC
WFDC2_79476_S	AAGAGCTCGTCAGGAGGGAGGCGTGGAC	AAGGATCCACTCACCTGAGACTAGGGTG
WFDC3_79509_S	AAAAGCTTTAAGTGTAACTGTCCTGGGC	AAGAGCTCCCTCACCGGGGGCGCCCTCA
WFDC3_79513_S	AAAAGCTTTACCTCTACCCTTTTCCAGG	AAGAGCTCAAGCCACCATTGCGGCACTG
ZNF313_79663_S	AAGAGCTCAGTGGTGCCGAACTAACGAT	AAGGATCCTGTACCGGCTTCTCGTACAC
ZNF334_79575_S	AAGGATCCCCTATAGGTTGTGGGTGCAC	AAGAGCTCCGGTGAAACGGATATGAAAC

B2. Primers for Amplification of 600 bp constructs (see section 4.3.1)

Primer Name (20ace database)	**Primer1**	**Primer2**
C20ORF100_79329	AAGCTAGCATATCTCCTCCTCCCTAGTG	AAGGATCCCCCCATCCCATTCGTATAAG
C20ORF111_79334	AAGGATCCGCTGCAGCCAAAGCACAAAG	AAGCTAGCAGGTAAGTGCCCTGACGTGAC
C20ORF121_79370	AAGCTAGCGCTCCAGAGCAAGAATCAGG	AAGGATCCTCCCTCTCCTCAGCTCCCAG
C20ORF35_79441	AAGCTAGCAGTATCTCAGAGAGATCCCC	AAGGATCCTTGATGCTCTGCGCCTTTTG
KIAA1404_79649	AAGGATCCATGCCCAGTACATGGAGAAC	AAGCTAGCGCACGTTCATTGTCGTTCTG
LPIN3_80393	AAGCTAGCTTACAACTGTGGGCTATGCG	AAGGATCCAGAGGGTCATCTCTTCTCCC
MATN4_80334	AAGGATCCAGCAAAGAAGGCCTCTCATG	AAGCTAGCAGAGCATTTAACACTGTGCC
PRKCBP1_79596	AAGGATCCATGGAATGCTGTCATGAGCC	AAGCTAGCTGGAAGGTAAGCAAACAGGC
SLPI_80494	AAGGATCCAGTGACTCTGATGGCCAATG	AAGCTAGCGGAGCTCTTCTTCAGCTTTC
UBE2V1_80531	AAGGATCCTACTCGCATGGATTCCTTGG	AAGCTAGCGTAAAAAGGCAAACCTGCCC
CSE1L_80345	AAGAGCTCAGAGGAACAGGAAGAAGGTG	AAGGATCCTTCAAACCCCGGCAAAATGG
R3HDML_79344	AAGAGCTCTGGTCTCTCGTGCTTCTCTG	AAGGATCCCTGAGCGTTCTCCAACAGTG
TDE1_80544	AAGGATCCTCCTGAGGCTGCTTTCTAAC	AAGAGCTCGAAATGGGACGTTCTCACTC
TOP1_80397	AAGAGCTCGTGGGCGTGAAATAATCCAG	AAGGATCCAGACTCCAGAAACGGCTGAG
YWHAB_79386	AAGAGCTCTTTCTGTTCTTCCCTGGCTC	AAGGATCCTGACCTCTTCCTTCCTAAGC
C20ORF167_79503	AAGAGCTCTGACAGAGTCCAGCGGAGTTG	AAAAGCTTAACCTGGACATCCCTGCTTC
ID1_long	CCGAGCTCAGGAGCTGCAAATTCAAG	AAGGATCCAGCCCGAAGCAGATAC
ID1_short	AAGAGCTCTTCCAGAGGAGCCCAG	AAGGATCCAGCCCGAAGCAGATAC

B3. Primers for Negative Control Fragments (300 bp) (see Chapter 4)

Primer Name (20ace database)	**Primer1**	**Primer2**
stSG116340	GCTATAGCACCATGCTGCAG	GGAGCAGAAGGGAAGGATCT

stSG116341	CCCTACCAAAAACAGAAAACA	GACAATGTCAGCTCCTGCTG
stSG116342	AATCCTCTGTTTCCCCGTTT	TGTTTGGTGAGTCTCTGGGA
stSG116343	AAACTGAAAATGCATTATTGGT	ACAATTAAAGTGAGGCAAGGG
stSG116344	TCCCATTTTCCCTTACTCCC	TATTTGTGGTGGCCAGGAAT

B4. Primers to check success of cloning to pDrive subcloning vector (see section 2.1.2 and 2.1.3)

Primer Name (20ace database)	Primer1	Primer2
pDriveinsertcheck	GTAAAACGACGGCCAGT	AACAGCTATGACCATG
stSG1163639	TACTAACATACGCTCTCCATC	TTCCATCTTCCAGCGGATAG

Primers for Real Time PCR (see section 5.2)

Primer Name (20ace database)	Primer1	Primer2
C20orf121_RT_01	GCTGGTCTCGAACTCCTGAACT	AAAACCCCCCAAAACAGCC
C20orf121_RT_02	CCACCACGCCCAACTAATTTT	GGAGTATGATGAGACCACCCCA
C20orf121_RT_03	CCGGCCTTACCTTACGTTTTAA	ACTGACTGATCCTCCTGCTTCC
C20orf121_RT_04	AGAGGTGACAACAGAAGCCAGA	GTTTGTAGAGGCCACCTGCAT
C20orf121_RT_05	GTCAAAACTCATGATTCCCAGG	AAGAGGAGTCCCAAGAAGATGG
C20orf121_RT_06	AGGTAAGGAAACGGAAGCACAA	CCCACTATGGCTGCTATCAGGA
C20orf121_RT_07	GGTTTGAAGCCCTTTACTGCCT	CTGCCATTTTCCAAGAACATGC
C20orf121_RT_08	ACCTCCTTCCTTCTCTGAGCCT	AGCTGCGGACCTGGAGTG
C20orf121_RT_09	TTCTCTGAGCCTCAGTTTCCCC	GAGCTCCGAGCTGCGGAC
C20orf121_RT_10	AGATGGCTTCCTGACCTCTCCT	TGGTAACGGTGAAGCCAAAGA
C20orf121_RT_11	CTTCACCGTTACCAATGGTGTC	TGGAGCTTCATGTTGACTCTCC
ADA_RT_01	GGCCTGATCATACAGCTGAGCT	TCTGCTTCTGCCCCTGACTTGA
ADA_RT_08	GTCTCTGCCGGCTCGGTG	CCCGGCCCGTTAAGAAGA
ADA_RT_07	GGCCACGCTCTTCTTAACGG	CAGGAAATGCGCGATCCA
ADA_RT_09	TGGAATTCTGGACCCGGCGT	AAGGCAGCGCCCAGCGAG
ADA_RT_02	CACACCAGCGCTATTCCGAATA	TGTCTGGCTGAAGTTCATTCCG
ADA_RT_03	CGGAATGAACTTCAGCCAGACA	CCATTGTATGCAGTTTTCCGCA

ADA_RT_04	TGTTTAGACACATGCATCGGTG	ACCCGCGTGTTATTTCCCT
ADA_RT_05	GTGCCCTGCTAAGTTTTGGATT	TCATGTGAGGTCAGGTGTTCG
ADA_RT_06	GCTTCCCAAGGCGTGATTA	AGCTACCATCCACCCCAGTACT
ZNF335_RT_01	GCCTAACTCTACCCGAAGCTCA	GAGAGGAACGTGGCTACGAAA
ZNF335_RT_02	AGGAACCCATCGGCCTATTGT	TCCGGAACACTGAAAATGTGTC
ZNF335_RT_03	TTTGGCACTGAGCCCACTATG	CGTAACACTCCGCCGAGATAAA
ZNF335_RT_04	TGCAACCCAATCGAGCTCT	CCCATACCCAGACCTCAGTTTC
ZNF335_RT_05	GCTGGTCCAAAATCACATCCA	GCCATCCGTACTTAGCTCCTGA
ZNF335_RT_06	CTTGATCCCAGGAGTTCCAAAC	CACCATGCCTGGCTAATTTTTC
ZNF335_RT_07	AAAATTAGCCAGGCATGGTGG	AGCCTCAACCTCCCAGACTCAA
c20orf142_RT	CCCTGGAGAATTGGAGCAGTAG	TCCGAGGCTCCAAGAAAGATC
YWHAB_RT	AACCCTGCCTTTCGCTCTT	CCCAAACCGTGTTCTCATTG
TOMM34_RT	CGAATATCTCCGCCCAACA	GCCTAAGTTCCAAACCGAGCA
SLPI_RT	GCCAATGCCAAACCTCACTATT	GGATTTCTGGTCTCTGGCTACG
PIGT_RT	AGCGGAAGTCACTTCTCACACG	CCATGAAAAGCCGGTATCCAAT
H3K27me3_Pr11	ACCACGCCCGGCTAATTTT	GGCGGATCATGAGGTCAGG
H3K27me3_Pr12	CTGACCTCATGATCCGCCC	GCCAGGTACAGTGGCTCACG
H3K27me3_Pr13	TGCAGTCTCCATGTGGCAAG	CCTGCCAGCCTTAACTCCATC
H3K27me3_Pr14	GATGGAGTTAAGGCTGGCAGG	CCACATCCAGGCATCCCTC
H3K27me3_Pr15	CTGCTGTGTGACCCCTGGTTA	CCGCTCTGCTCCAAGCACT
H3K27me3_Pr16	GAAGAAAGTGCCCTGCCCA	TACAGGCGTGAGCCACTGC
H3K27me3_Pr17	GTCAGGAATTCGAGACCAGCC	CACCCACCATCACACCCAA
H3K27me3_Pr21	CTGAAAGTGACGACACAGCCA	CCTGAGGTGAAAGGATGTTTCC
H3K27me3_Pr22	CATAACTCCAGCCCCCTTTTTT	CCTGTTCTCAGCACATCATGGT
H3K27me3_Pr23	CCAAAGACTTTTCCTGAGCACC	GGTGCAATCGTCAGCTCTTTCT
H3K27me3_Pr24	CACTTGGAATTTGATCAGCGG	GGCCCCATAGGTACATTTCCTT
H3K27me3_Pr25	CCCTGGTGCCCAAGATACTTTC	AGCAAAGAGGCTAGGCAGGAGA
H3K27me3_Pr26	ACTTGAGTGCCCTACAGTGCCT	GAATGGTTGGATGATGGTGCTA
H3K27me3_Pr27	ATCCTAGGTTCTTCCTCTCCCC	CCTGACCAGAGCAATGTAGAGC
CTCF_Pr11	CCCTCACGGTTCCTATTCTGC	TGACTCACCACGCCCCATA
CTCF_Pr12	CGACACGTCACTCAATCGCTT	CAGTAACCCAGCGAGGTGGA
CTCF_Pr13	CTCACACACAAGCGCCGTG	GGCCAAGGTGACCAGTTGG
CTCF_Pr14	CAACTGGTCACCTTGGCCA	GCCATGCCAACCA CTTACCT
CTCF_Pr15	ACCCTTGGAGACCCCACGT	GGTGAATCGGTGAGGACGG
CTCF_Pr16	GATTTCCCGTCCATATCCGC	GACTGAGTCCCCGATCTCCTG

CTCF_Pr21	ATCGGCAAGAACACAAGACTGG	CCTGGCCTTGTCGTTCGTC
CTCF_Pr22	GACGAACGACAAGGCCAGG	CCTCTACTGGCCAAGTCGGAG
CTCF_Pr23	GGCCAGTAGAGGCAGTGAGGA	TGAGCCCTATGTCTGGCCAG
CTCF_Pr24	ATCAGAGTCATGGCACCCCA	CAGCCCATCACCTACTCTGCA
ChIP_nctrl01	TCAGGCTTTTTCACTGCCTT	GTGTATGGGGAATGGAGGTG
ChIP_nctrl02	TTGAGCATCTGCTATGTGCC	GCTGTACTTCAGCCTGGGAG
ChIP_nctrl03	TCACTTTGATGCTTGGCTTG	ATCCCCAGAACCTGTGAGTG
ChIP_nctrl04	GCGGTCTTTGTAAAACCCAA	AGTTCTGCCTGGACTCCTGA
ChIP_nctrl05	TAACGTGCCACCTACTTCCC	GTGCACCAGGCCTTTAACAT
ChIP_nctrl06	TCACCTGAGGTCAGGAGCTT	CTTGCTCAAGGTTTCAAGGC
ChIP_nctrl07	ATAGTCGGTGGCCAACAAAG	TGATCTCTCCAGCCTCAGGT
ChIP_nctrl08	CTTGAGCCAGTCCCTTTCTG	TCTAGTGCAAGGCCACCTCT
ChIP_nctrl09	GACTGTGGTGATGGTTGCAC	CAGGTGGTTTCAGGTGGTCT
ChIP_nctrl10	GTGATTCTTCTGCCTCAGCC	ACTGGAAACACCATTGGCTC
ChIP_nctrl11	AGAGGGGCATTGTGGTGTAG	TGAGGCGATTCAACTCTGTG
ChIP_nctrl12	CTACCGCTTGATGGCTTCTC	TTCCCTCTGTGTGCTGAGTG
ChIP_nctrl13	ACCCTGAACCTGCCATACTG	CATCACAGCAAGCCTTTGAA
ChIP_nctrl14	AAATCCAGACACCTGGCAAC	GTCAGGAGTTCGAGACCAGC
ChIP_nctrl15	TGTGGTAGTGTGCCCCTGTA	TTCAACTCTTTGCCAGCCTT
ChIP_nctrl16	ACCACTTGAGTCTTGTGGGG	TCCTGTGATTGTTCAGCAGC
ChIP_nctrl17	CTGAAACCAAGCAGAAAGCC	GTCAGGCTGGTCTCGAACTC
ChIP_nctrl18	GAGCTGGGATTCGTGGATAA	AAGTGCCCTGGTGACATAGG
ChIP_nctrl19	CTCGTAGCCTCAAGCAATCC	CACCATGCCTGGCTAATTTT
ChIP_nctrl20	ATTACAGGCATGCAACACCA	CTTGCCTCCAAGGAACTCAG
chiparray_nctrl21	CATCCATTCAGTCATGTCGC	CACACTGTTTCCCCTCCACT
chiparray_nctrl22	TGCAGTAGGAACCCAGCTCT	ATGAGCGTGCCCATAAAAAC
chiparray_nctrl23	GCTGCATTGTAAACCACCCT	GGTAGACTGCTTGAGCCCAG
chiparray_nctrl24	CTTCCAGGGAGAAAGCACAG	ATTGTTCTGGCCAAAATTGC
chiparray_nctrl25	AACAAAATGTACCTTGCGGC	CTGACTCTGCACCAAGTGGA

B5. Primers to amplify fragments printed onto the 3.5 Mb custom-made array (see section 5.2)

	aminolinked sense	antisense
pUC18	**CCAGTCACGACGTTGTAA**	**CGGATAACAATTTCACACA**

B6. Primers Designed to fill gaps containing TSSs in 3.5 Mb custom made array (see section 5.3)

Primer Name (20ace database)	Primer1	Primer2
C20ORF146	CCAGTCACGACGTTGTAACATAAAACATTTTGTAATTAG	CCTTTCACTATGGGATGTGCCATTTC
NCOA5	CCAGTCACGACGTTGTAACATTTGGTAAATAGCAGTAC	TTAAATGAGATGATATGTGTAAAGGATTC
UBE2C	CCAGTCACGACGTTGTAACCTGAACCAATGTTGCCTTT	CAGTCACAGGAGGAGTGTCTTGCTCC
C20ORF170	CCAGTCACGACGTTGTAACTTTACCTTGCTCAGGCTCT	GACTCCATTGATATAGAACAGAAGC
PPGB	CCAGTCACGACGTTGTAAGAAGCAAACACGTCCACCAC	GAAATGCCTTACAAGGTCGCC
NCOA3	CCAGTCACGACGTTGTAAGTCATAAATACGCTGTGGGA	GATCTGAAGCCGCTGGCTCTCGCAG
OVCOV1	CCAGTCACGACGTTGTAATCGAAGGCTTAGTCTCTATT	CTTCTGGATCGCCCAAGTTTTAAG
TOMM34	CCAGTCACGACGTTGTAATGCCGGAGTCGGAAGGGGCT	CTCAGCATGTTGGCTTTTCATACTTG