US 20150307937A1

(19) **United States**

(12) **Patent Application Publication**

**Lipovich et al.**

(10) **Pub. No.: US 2015/0307937 A1**

(43) **Pub. Date: Oct. 29, 2015**

(54) **ACTIVITY-DEPENDENT GENE PAIRS AS THERAPEUTIC TARGETS AND METHODS AND DEVICES TO IDENTIFY THE SAME**

(71) Applicants: **Leonard LIPOVICH**, Detroit, MI (US); **Jeffrey A. LOEB**, Beverly Hills, MI (US)

(72) Inventors: **Leonard Lipovich**, Detroit, MI (US); **Jeffrey A Loeb**, Beverly Hills, MI (US)

(73) Assignee: **WAYNE STATE UNIVERSITY**, Detroit, MI (US)

(21) Appl. No.: **14/423,086**

(22) PCT Filed: **Aug. 22, 2013**

(86) PCT No.: **PCT/US13/56242**

§ 371 (c)(1),
(2) Date: **Feb. 20, 2015**

**Related U.S. Application Data**

(60) Provisional application No. 61/692,162, filed on Aug. 22, 2012.

**Publication Classification**

(51) **Int. Cl.**
***C12Q 1/68*** (2006.01)
***C12N 15/113*** (2006.01)

(52) **U.S. Cl.**
CPC ..... ***C12Q 1/6876*** (2013.01); ***C12N 15/113*** (2013.01); *C12N 2310/14* (2013.01); *C12Q 2600/158* (2013.01); *C12Q 2600/136* (2013.01)

(57) **ABSTRACT**

Activity-dependent gene pairs as therapeutic targets and methods and devices to identify the same are provided. The methods and devices allow transcriptome-wide analysis of regulatory long non-coding RNAs (lncRNAs), matched with differentially expressed protein-coding genes (mRNAs) and/or with other lncRNAs. The described methods and devices allow analysis of these activity-dependent gene pairs as therapeutic targets in a number of clinical conditions associated with altered electrical brain activity, including epilepsy.

| Patient | Gender | Age (years) | Interictal Spiking (High:Low) | Brain locations (High:Low) | Pathological Diagnosis |
|---|---|---|---|---|---|
| 1 | F | 15 | 6:0 | Left frontal/frontal | Polymicrogyria |
| 2 | F | 10 | 116:1 | Right temporal/temporal | Diffuse gliosis with inflammation |
| 3 | M | 33 | 371:115 | Right temporal/frontal | None |
| 4 | M | 27 | 17:3 | Left temporal/frontal | White matter gliosis |
| 5 | F | 3 | 178:58 | Left frontoparietal/frontoparietal | Mild gliosis |
| 6 | F | 7 | 215:25 | Left frontoparietal/frontoparietal | Cortical dysplasia |
| 7 | F | 6 | 124:36 | Right frontal/frontoparietal | Mild gliosis |

FIG. 1A

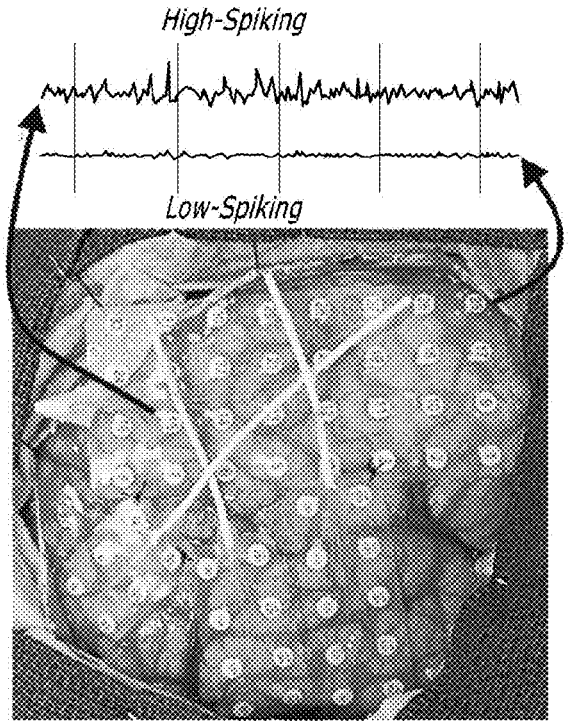

High-Spiking
Low-Spiking

FIG. 1B

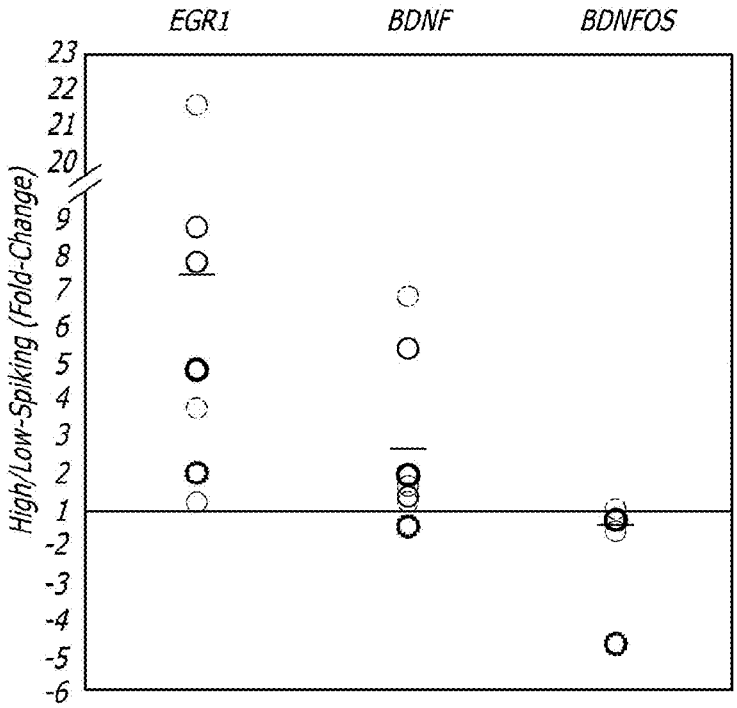

EGR1
BDNF
BDNFOS
High/Low-Spiking (Fold-Change)
23
22
21
20
9
8
7
6
5
4
3
2
1
-2
-3
-4
-5
-6

FIG. 1C

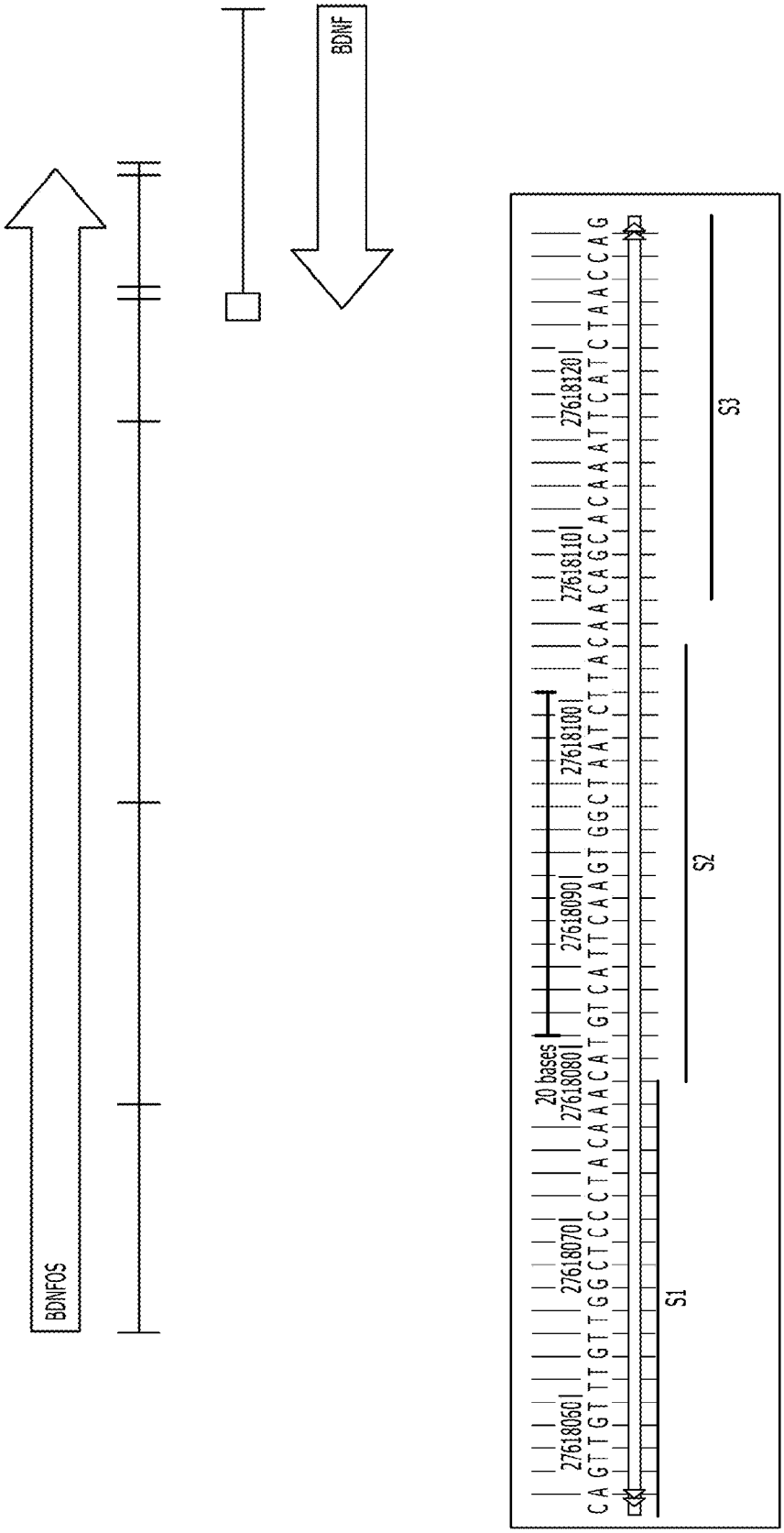

BDNF
BDNFOS
S1
S2
S3
20 bases
27618060
27618070
27618080
27618090
27618100
27618110
27618120
CAGTTGTTTGTTGGCTCCCTACAACATGTCATTCAAGTGGCTAATCTTACAACAGCACAATTCATCTAACCAG

FIG. 2A

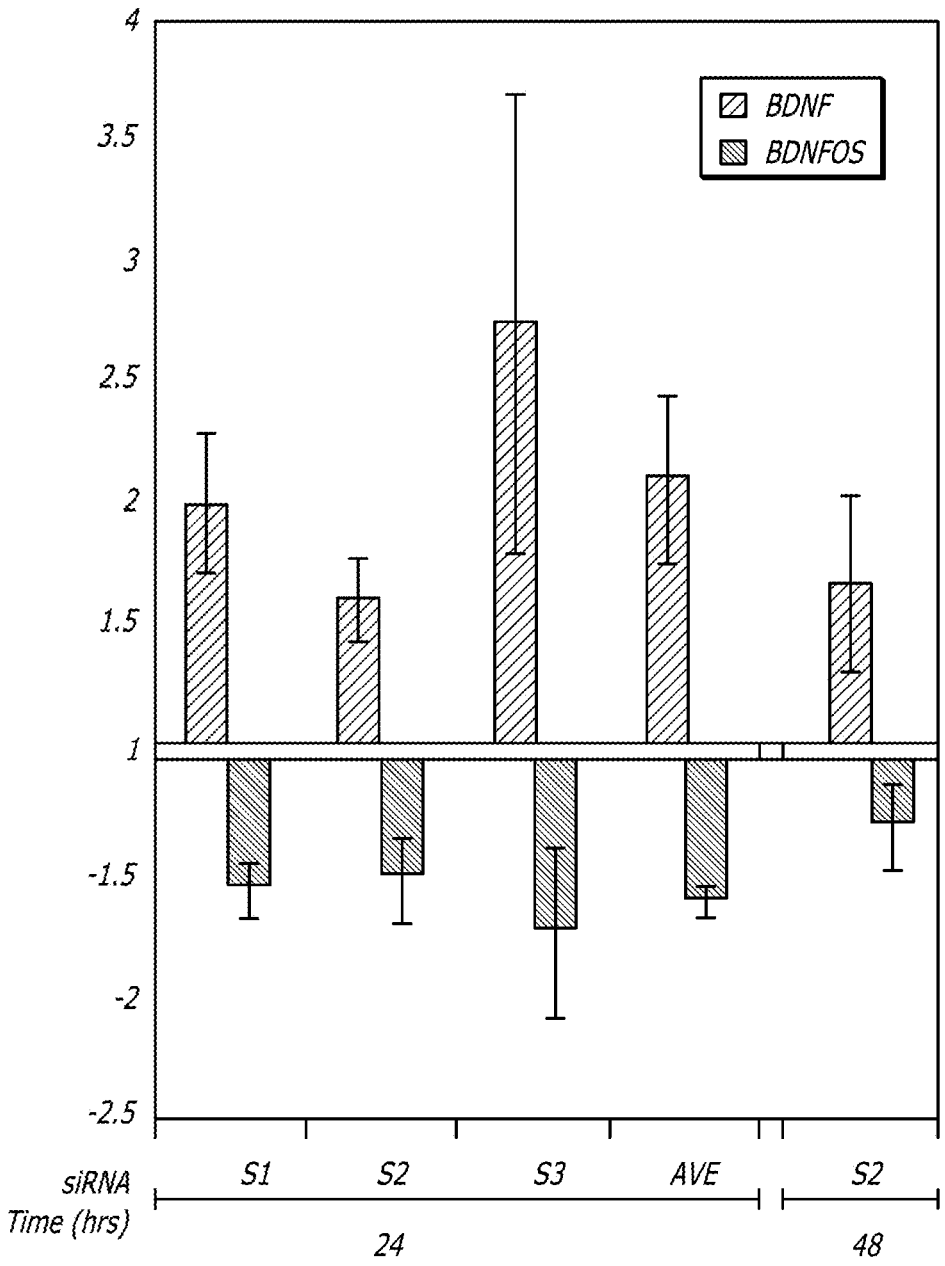

4
3.5
3
2.5
2
1.5
1
-1.5
-2
-2.5
BDNF
BDNFOS
siRNA
S1
S2
S3
AVE
S2
Time (hrs)
24
48

FIG. 2B

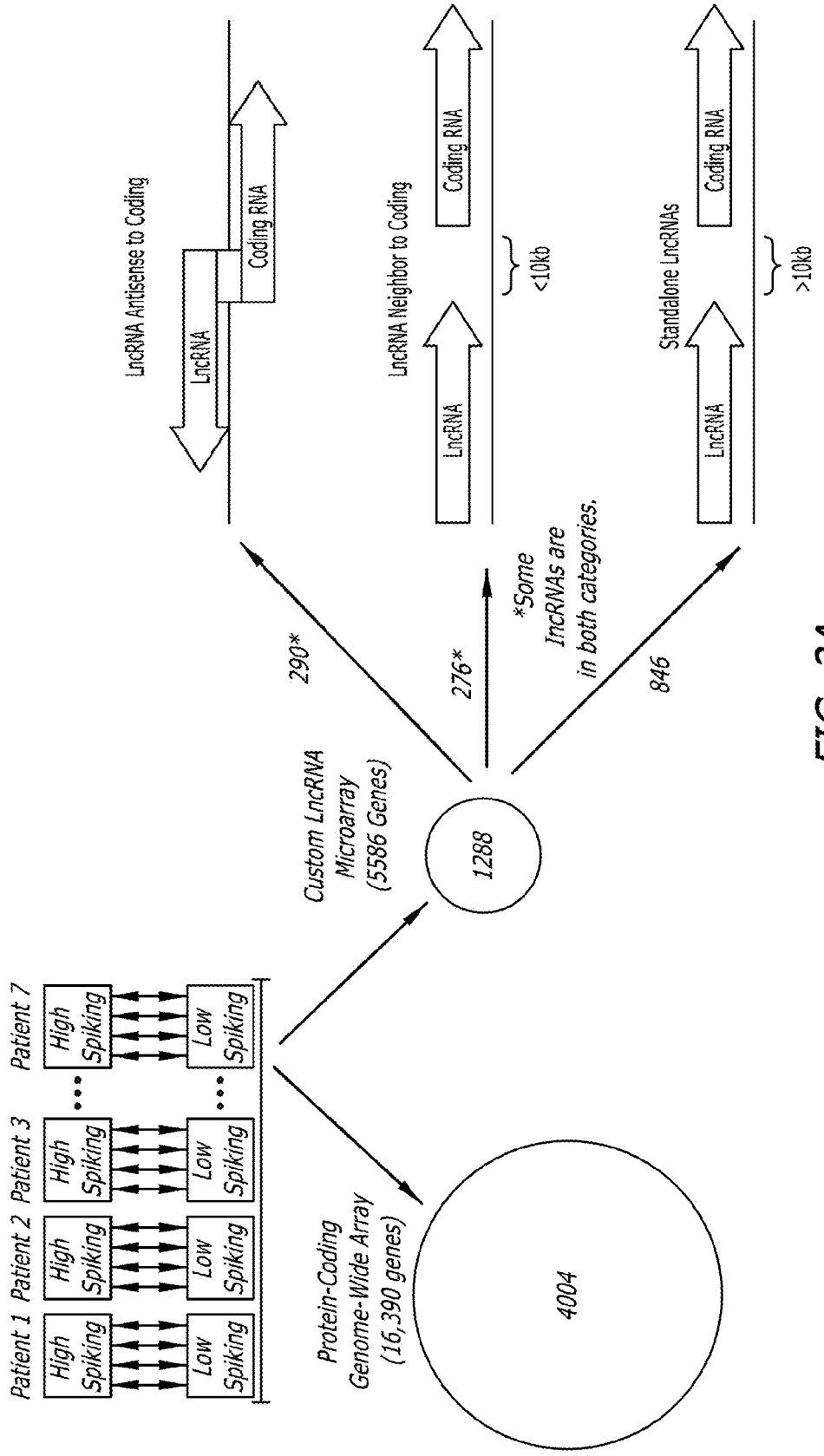
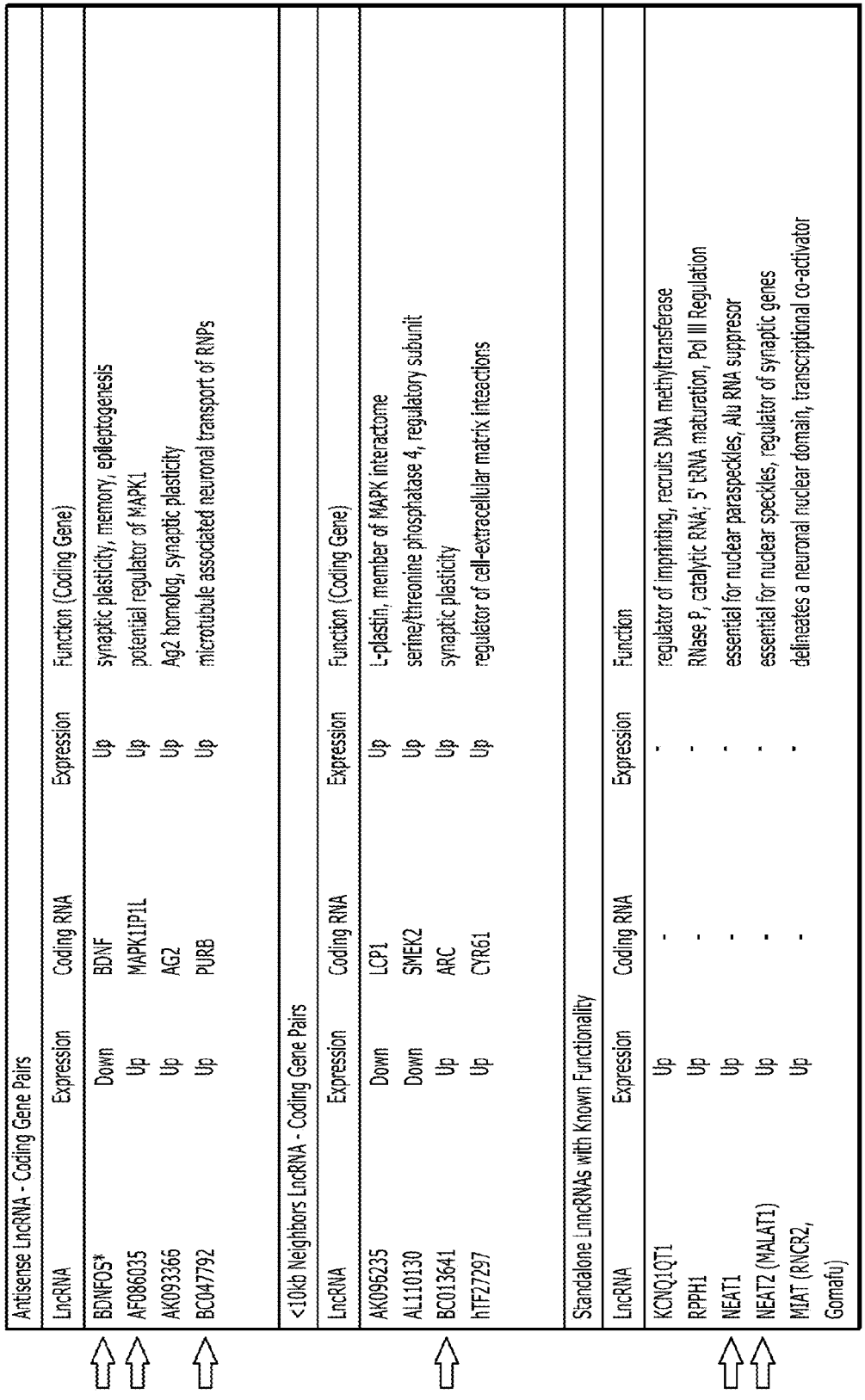

Patient 1
Patient 2
Patient 3
Patient 7
High Spiking
Low Spiking
Protein-Coding Genome-Wide Array (16,390 genes)
4004
Custom LncRNA Microarray (5586 Genes)
1288
290*
276*
846
*Some lncRNAs are in both categories.
LncRNA Antisense to Coding
LncRNA
Coding RNA
LncRNA Neighbor to Coding
<10kb
Standalone LncRNAs
>10kb

FIG. 3A



| Antisense LncRNA - Coding Gene Pairs | | | | |
|---|---|---|---|---|
| LncRNA | Expression | Coding RNA | Expression | Function (Coding Gene) |
| BDNFOS* | Down | BDNF | Up | synaptic plasticity, memory, epileptogenesis |
| AF086035 | Up | MAPK1IP1L | Up | potential regulator of MAPK1 |
| AK093366 | Up | AG2 | Up | Ag2 homolog, synaptic plasticity |
| BC047792 | Up | PURB | Up | microtubule associated neuronal transport of RNPs |

| <10kb Neighbors LncRNA - Coding Gene Pairs | | | | |
|---|---|---|---|---|
| LncRNA | Expression | Coding RNA | Expression | Function (Coding Gene) |
| AK096235 | Down | LCP1 | Up | L-plastin, member of MAPK interactome |
| AL110130 | Down | SMEK2 | Up | serine/threonine phosphatase 4, regulatory subunit |
| BC013641 | Up | ARC | Up | synaptic plasticity |
| hTF27297 | Up | CYR61 | Up | regulator of cell-extracellular matrix inteactions |

| Standalone LincRNAs with Known Functionality | | | | |
|---|---|---|---|---|
| LncRNA | Expression | Coding RNA | Expression | Function |
| KCNQ1QT1 | Up | - | - | regulator of imprinting, recruits DNA methyltransferase |
| RPPH1 | Up | - | - | RNase P, catalytic RNA; 5' tRNA maturation, Pol III Regulation |
| NEAT1 | Up | - | - | essential for nuclear parapeckles, Alu RNA suppresor |
| NEAT2 (MALAT1) | Up | - | - | essential for nuclear speckles, regulator of synaptic genes |
| MIAT (RNCR2, Gomafu) | Up | - | - | delineates a neuronal nuclear domain, transcriptional co-activator |

FIG. 3B

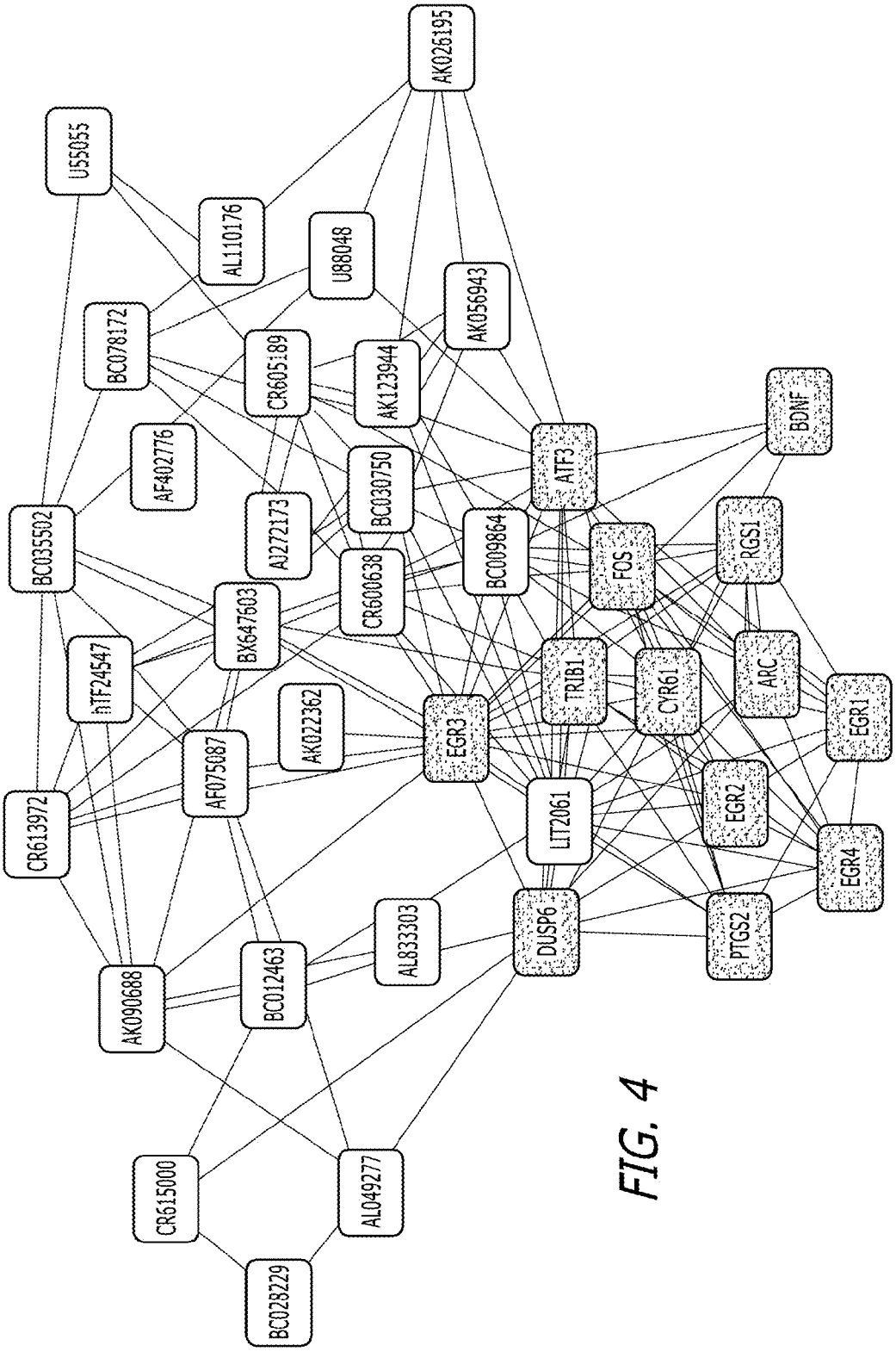
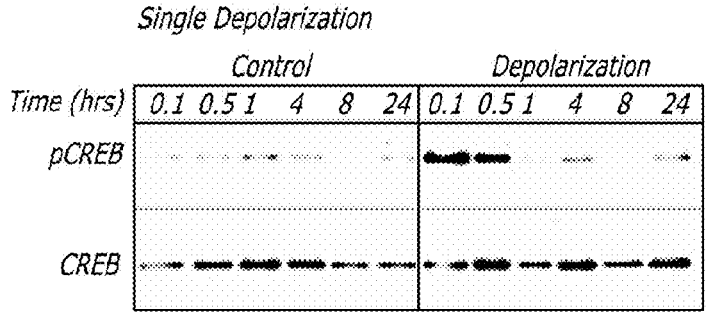
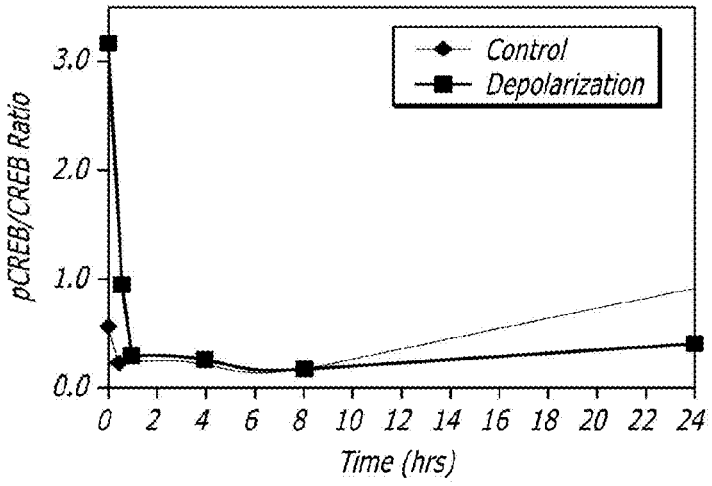

AK026195
U55055
AL110176
U88048
AK056943
BC078172
CR605189
AK123944
BDNF
AF402776
ATF3
BC030750
BC035502
AJ272173
BC009864
FOS
RGS1
CR600638
BX647603
hTF24547
TRIB1
CYR61
ARC
AK022362
EGR3
EGR1
AF075087
CR613972
LIT2061
EGR2
EGR4
AL833303
DUSP6
PTGS2
BC012463
AK090688
CR615000
AL049277
BC028229

FIG. 4

FIG. 5A


Single Depolarization
Control
Depolarization
Time (hrs)
0.1 0.5 1 4 8 24
0.1 0.5 1 4 8 24
pCREB
CREB


Control
Depolarization
pCREB/CREB Ratio
3.0
2.0
1.0
0.0
0 2 4 6 8 10 12 14 16 18 20 22 24
Time (hrs)

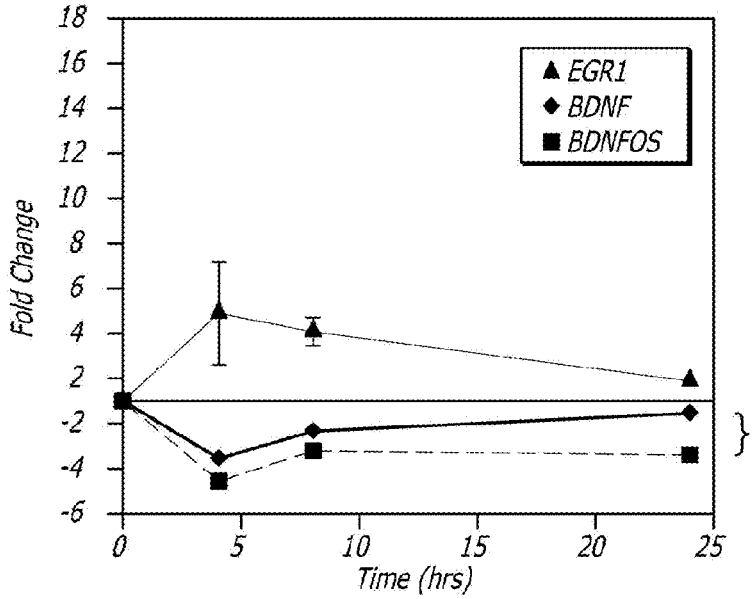
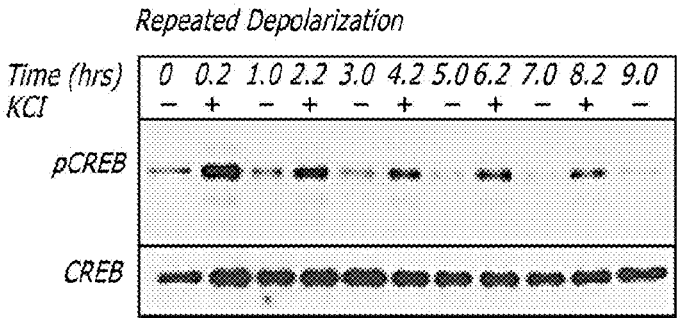
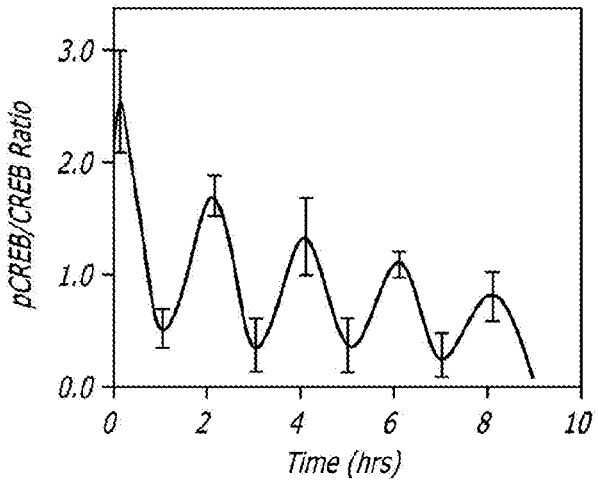

EGR1
BDNF
BDNFOS
Fold Change
18
16
14
12
10
8
6
4
2
-2
-4
-6
0 5 10 15 20 25
Time (hrs)

FIG. 5B


Repeated Depolarization
Time (hrs)
KCl
0 0.2 1.0 2.2 3.0 4.2 5.0 6.2 7.0 8.2 9.0
− + − + − + − + − + −
pCREB
CREB


pCREB/CREB Ratio
3.0
2.0
1.0
0.0
0
2
4
6
8
10
Time (hrs)

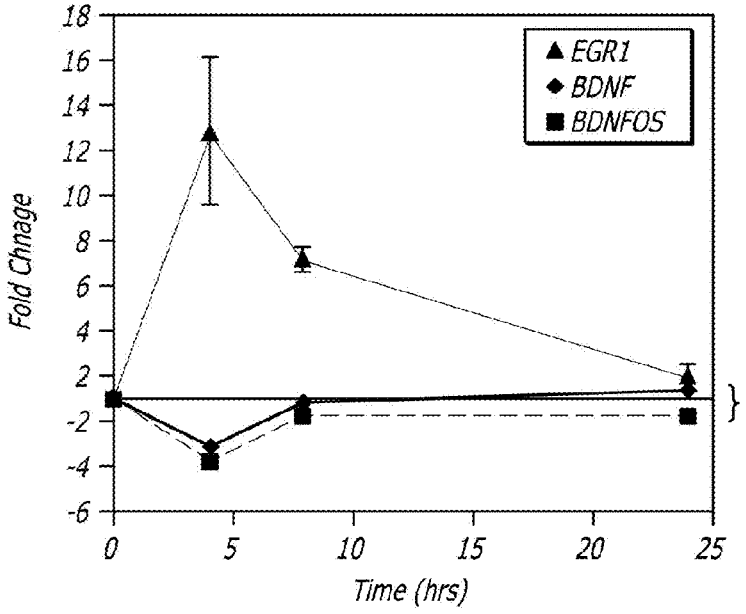
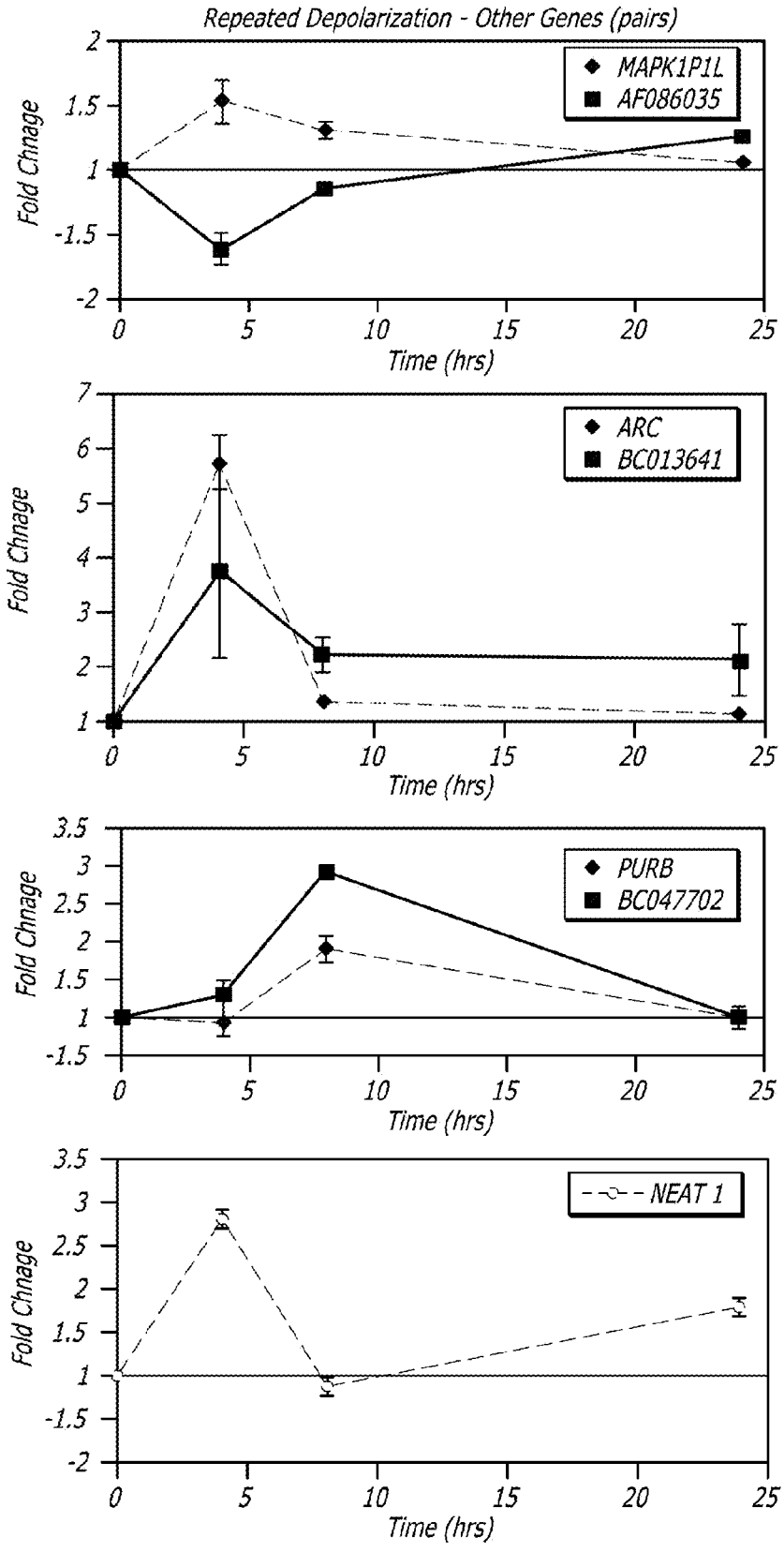

Fold Chnage
18
16
14
12
10
8
6
4
2
-2
-4
-6
0
5
10
15
20
25
Time (hrs)
EGR1
BDNF
BDNFOS

FIG. 5C


Repeated Depolarization - Other Genes (pairs)
MAPK1P1L
AF086035
Fold Chnage
Time (hrs)
ARC
BC013641
Fold Chnage
Time (hrs)
PURB
BC047702
Fold Chnage
Time (hrs)
NEAT 1
Fold Chnage
Time (hrs)

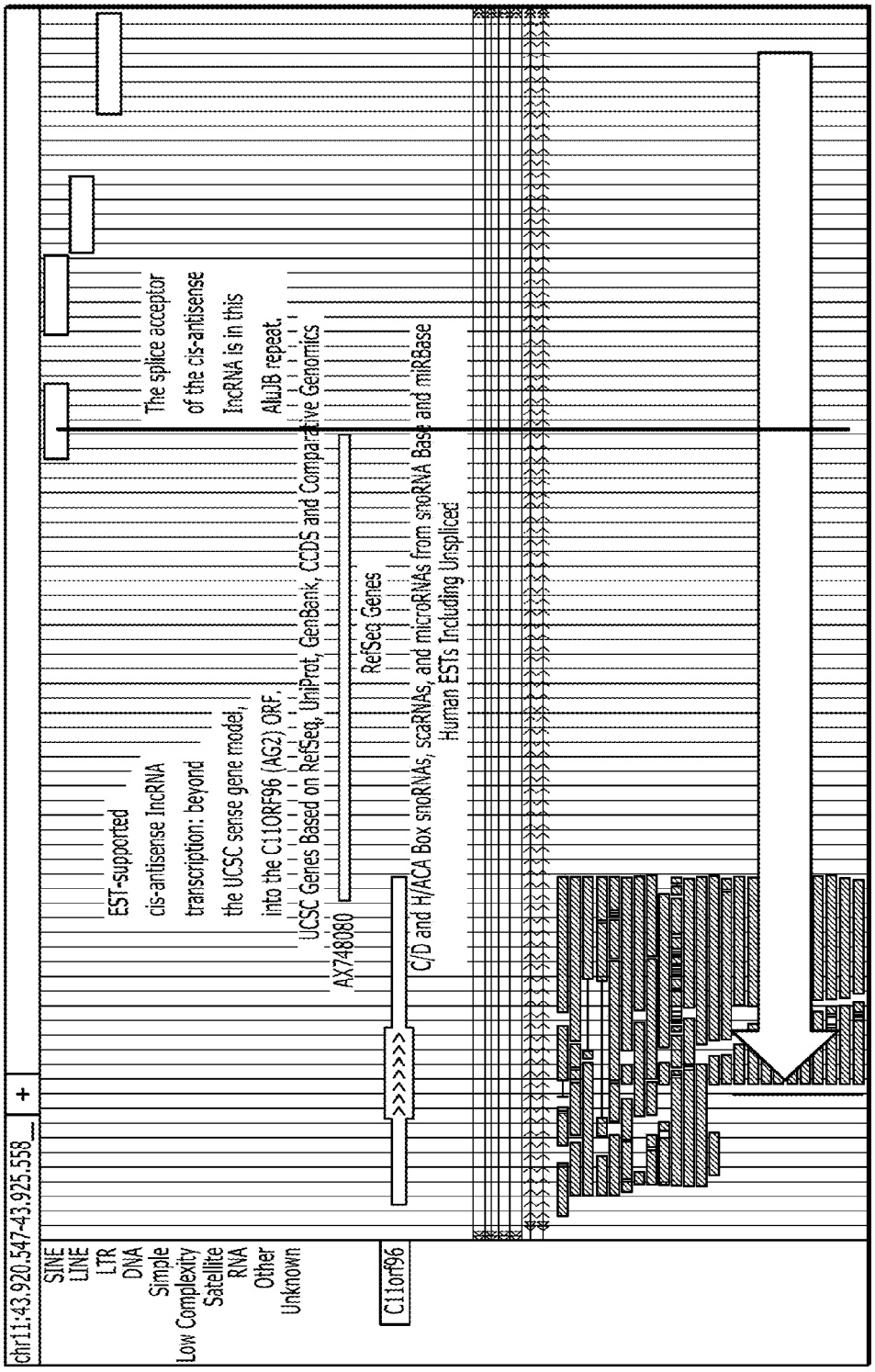

chr11:43,920,547-43,925,558
+
SINE
LINE
LTR
DNA
Simple
Low Complexity
Satellite
RNA
Other
Unknown
C11orf96
EST-supported
cis-antisense lncRNA
transcription: beyond
the UCSC sense gene model,
into the C11ORF96 (AG2) ORF,
AX748080
UCSC Genes Based on RefSeq, UniProt, GenBank, CCDS and Comparative Genomics
RefSeq Genes
C/D and H/ACA Box snoRNAs, scaRNAs, and microRNAs from snoRNA Base and miRBase
Human ESTs Including Unspliced
The splice acceptor
of the cis-antisense
lncRNA is in this
AluJB repeat.

FIG. 6

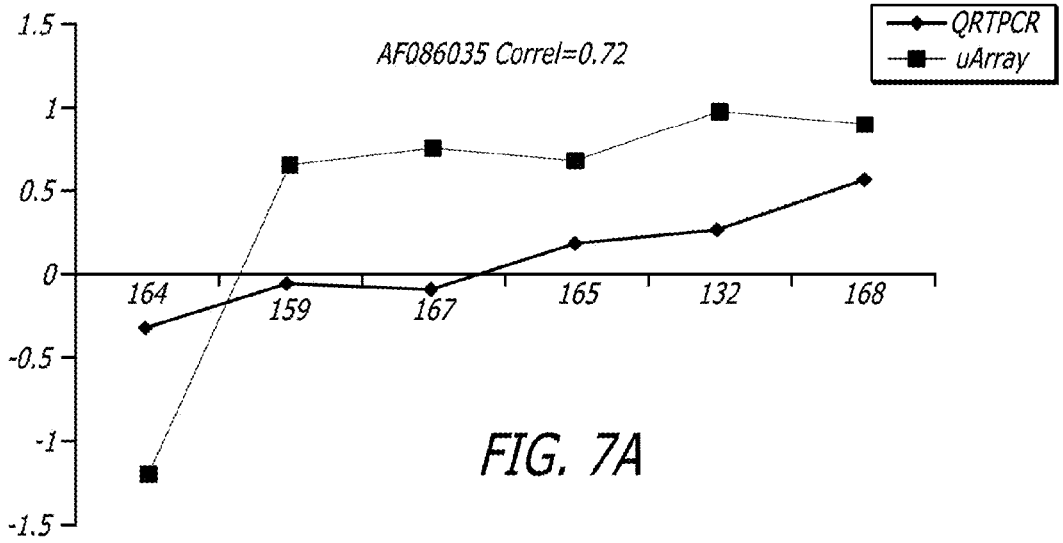

AF086035 Correl=0.72
QRTPCR
uArray
1.5
1
0.5
0
-0.5
-1
-1.5
164
159
167
165
132
168

FIG. 7A

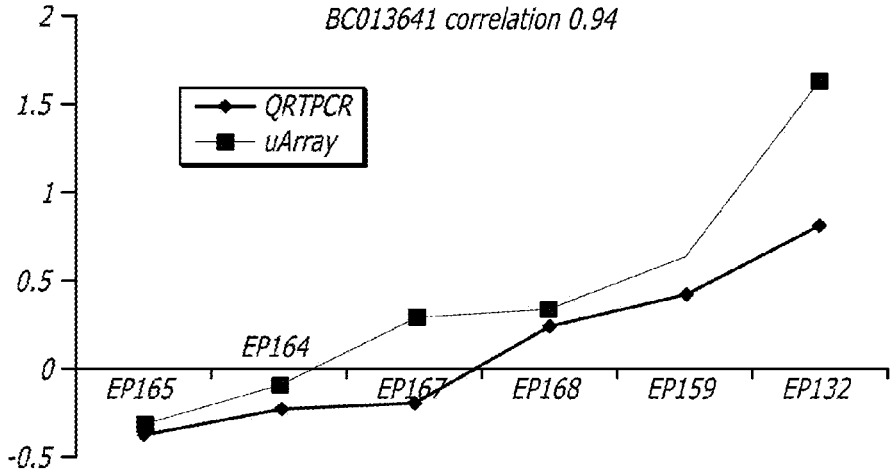

BC013641 correlation 0.94
QRTPCR
uArray
2
1.5
1
0.5
0
-0.5
EP164
EP165
EP167
EP168
EP159
EP132

FIG. 7B

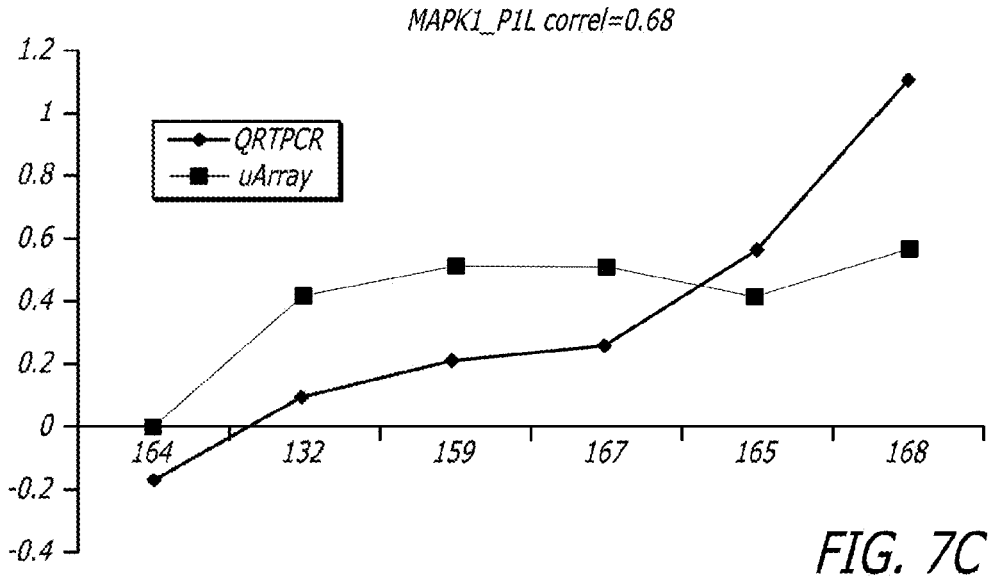

MAPK1_P1L correl=0.68
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
QRTPCR
uArray
164
132
159
167
165
168

FIG. 7C

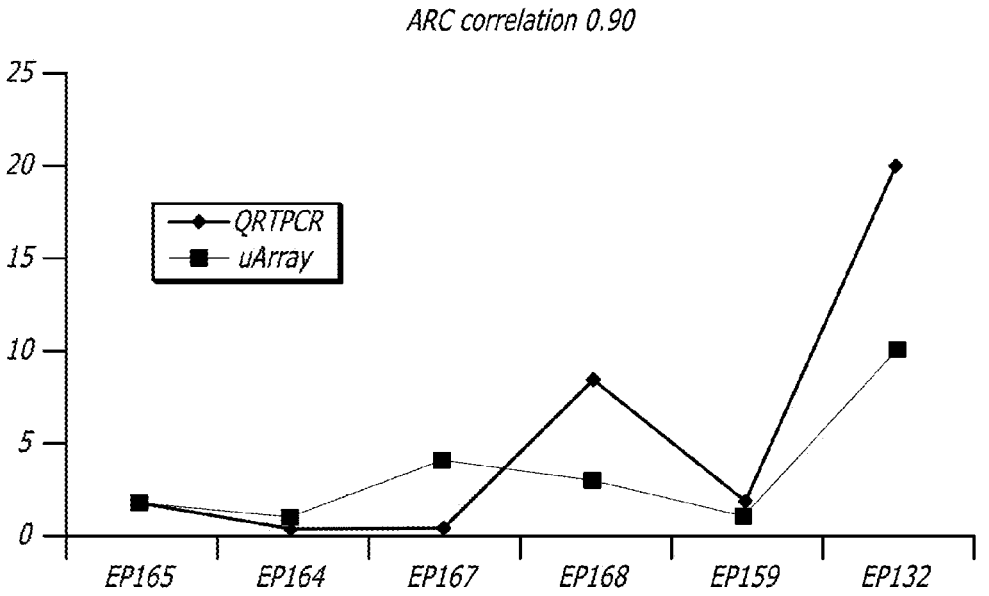

ARC correlation 0.90
25
20
15
10
5
0
QRTPCR
uArray
EP165
EP164
EP167
EP168
EP159
EP132

FIG. 7D

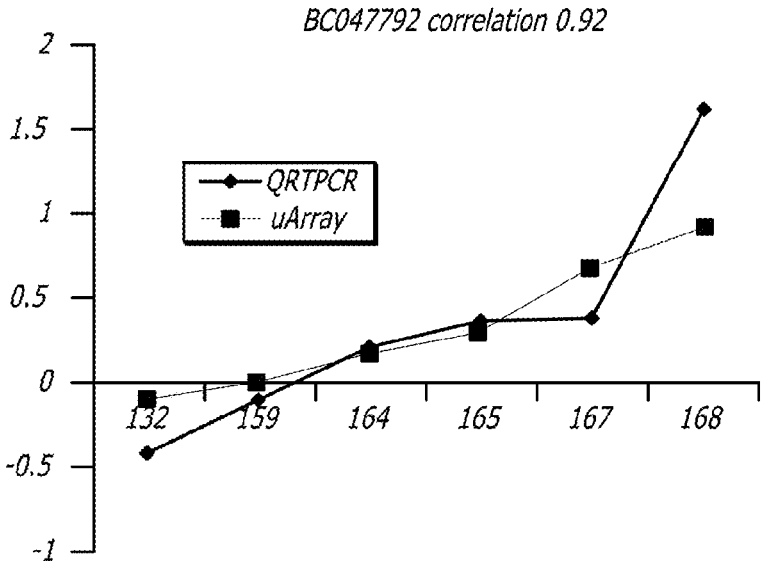



FIG. 7E

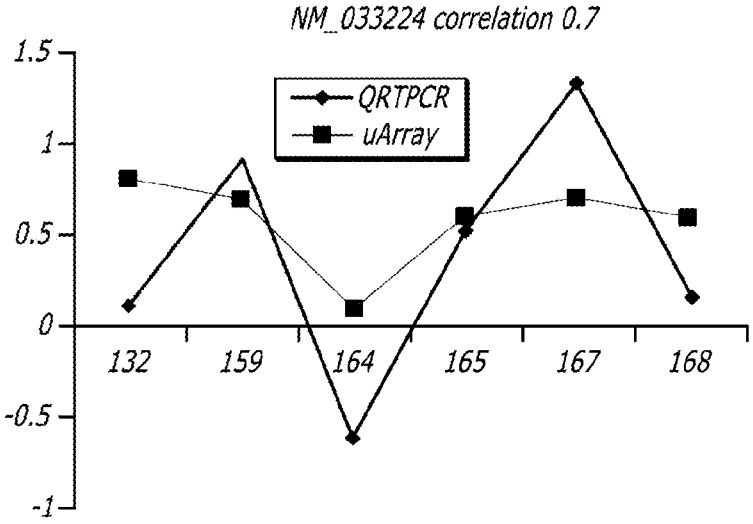



FIG. 7F

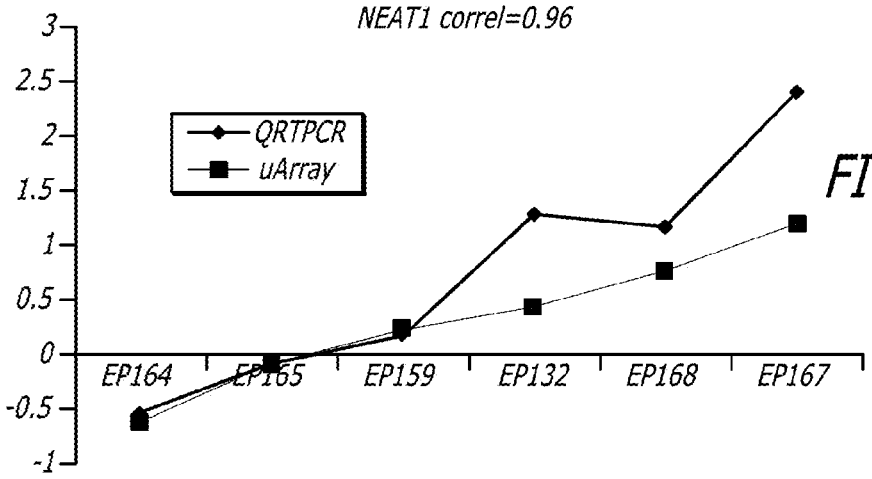



FIG. 7G

# ACTIVITY-DEPENDENT GENE PAIRS AS THERAPEUTIC TARGETS AND METHODS AND DEVICES TO IDENTIFY THE SAME

## STATEMENT OF GOVERNMENT INTEREST

**[0001]** This work was funded in part by NIH/NINDS R01NS045207, R01NS058802 and NIH/NIDA 1R03DA026021-01.

## FIELD OF THE DISCLOSURE

**[0002]** Disclosed herein are activity-dependent gene pairs as therapeutic targets and methods and devices to identify the same. The methods and devices allow transcriptome-wide analysis of regulatory long non-coding RNAs (lncRNAs), matched with differentially expressed protein-coding genes (mRNAs) and/or with other lncRNAs. The described methods and devices allow analysis of these activity-dependent gene pairs as therapeutic targets in a number of clinical conditions associated with altered electrical brain activity, including epilepsy.

## BACKGROUND OF THE DISCLOSURE

**[0003]** Many of the body's functions are mediated by the actions of proteins within and between cells. Generally, for a protein to exert an effect, the cell that will use or secrete the protein must create it. To create a protein the cell first makes a copy of the protein's gene sequence in the nucleus of the cell. This copy of the gene sequence that encodes for the protein (called messenger RNA (mRNA)) leaves the nucleus and moves to a region of the cell containing ribosomes. Ribosomes read the sequence of the mRNA and create the protein for which it encodes. This process of new protein synthesis is known as translation.

**[0004]** Long non-coding RNAs (lncRNAs) are a newly discovered type of RNA that are generated by cells but do not encode any protein. Despite their new-found prominence in the transcriptome (the total set of all RNA molecules), most lncRNAs remain poorly understood.

## SUMMARY OF THE DISCLOSURE

**[0005]** The present disclosure provides activity-dependent gene pairs as therapeutic targets and methods and devices to identify the same.

**[0006]** Particularly, one embodiment includes a method for identifying putative therapeutic targets comprising obtaining a paired brain tissue sample from a live human wherein each member of the pair has a different level of electrical brain activity from the other member; identifying long non-protein-coding RNA (lncRNA) molecules (lncRNAs) and protein-coding messenger RNA (mRNA) molecules (mRNAs) that are differentially expressed between the members of each individual sample pair; linking a first differentially expressed lncRNA with a differentially expressed mRNA and/or a second differentially expressed lncRNA when the gene encoding the first differentially expressed lncRNA overlaps with, or is adjacent to, the gene encoding the differentially expressed mRNA and/or the gene encoding the differentially expressed second lncRNA along the human genome, thereby identifying an lncRNA/mRNA gene pair and/or an lncRNA/lncRNA gene pair as putative cis-encoded therapeutic targets; and/or linking a first differentially expressed lncRNA with a differentially expressed mRNA and/or with a second differentially expressed lncRNA when the differentially expressed first lncRNA and the differentially expressed mRNA and/or second lncRNA are encoded at different genomic loci, thereby identifying an lncRNA/mRNA gene pair and/or an lncRNA/lncRNA gene pair as putative trans-encoded therapeutic targets.

**[0007]** In another embodiment, the linking of differentially expressed lncRNA with differentially expressed mRNA and/or lncRNA further requires that the differential expression of the lncRNA and mRNA or lncRNA and lncRNA be observed in more than one brain sample pair, each pair having a low electrical brain activity member and a high electrical brain activity member.

**[0008]** In another embodiment, the electrical brain activity is classified as high or low based on the frequency and/or amplitude of interictal and/or ictal spiking.

**[0009]** In another embodiment, the differential expression is identified by quantifying lncRNA and mRNA expression. In another embodiment, the expression quantification utilizes at least one microarray capable of quantifying lncRNA expression and mRNA expression.

**[0010]** In another embodiment, the quantifying utilizes at least one microarray capable of quantifying lncRNA expression and at least one microarray capable of quantifying mRNA expression wherein consistency of differential expression data between the at least one lncRNA microarray and the at least one mRNA microarray is evaluated by correlating the fold-change of protein-coding control genes common to both arrays.

**[0011]** Another embodiment further comprises evaluating the putative therapeutic target as a site of effective intervention.

**[0012]** In another embodiment, the therapeutic target of a pair is lncRNA, mRNA and/or both.

**[0013]** Another embodiment includes a microarray for identifying putative therapeutic targets in the human brain comprising probes for lncRNA and probes for mRNA wherein at least a subset of the mRNA probes is included based on the representation of their corresponding genes by probes on a different genomewide expression analysis microarray.

**[0014]** In another embodiment, the lncRNA probes are 50-mer to 70-mer probes mapped to a single genomic location. In another embodiment, the lncRNA probes are free of interspersed and simple repeats and segmental duplications.

**[0015]** In another embodiment, the microarray comprises 7 or 8 distinct probes per lncRNA.

**[0016]** In another embodiment, the microarray comprises probes for at least 1000 lncRNA genes.

**[0017]** Another embodiment includes a method of assessing putative therapeutic targets in the human brain comprising exposing human neuroblastoma cells to either a single depolarization or repeated depolarizations; identifying time-dependent differential lncRNA and mRNA expression in the cells exposed to either single and/or repeated depolarizations, relative to untreated control cells; and (i) linking a first differentially expressed lncRNA with differentially expressed mRNAs and/or second differentially expressed lncRNAs when the gene encoding the first differentially expressed lncRNA overlaps with, or is adjacent to, the gene encoding the differentially expressed mRNA or second differentially expressed lncRNA thereby identifying lncRNA/mRNA and/or lncRNA/lncRNA gene pairs as putative cis-encoded therapeutic targets; and/or (ii) linking a first differentially expressed lncRNA with a differentially expressed mRNA

and/or with a second differentially expressed IncRNA when the first IncRNA and mRNA and/or second IncRNA are encoded at different genomic loci, thereby identifying IncRNA/mRNA and/or IncRNA/IncRNA gene pairs as putative trans-encoded therapeutic targets.

**[0018]** Another embodiment includes a method comprising targeting the first IncRNA of an IncRNA/mRNA or IncRNA/IncRNA gene pair as a putative therapeutic target wherein the first IncRNA and mRNA or second IncRNA are differentially expressed in areas of the brain having a different characteristic demonstrated to be relevant in one or more of epileptic activity, inflammation, cellular proliferation multiple sclerosis, neurodegeneration or autism and/or have been linked because the gene encoding the differentially expressed IncRNA overlaps with, or is adjacent to, the gene encoding the differentially expressed mRNA or second differentially expressed IncRNA.

**[0019]** In another embodiment, the gene pair is BDNFOS (SEQ ID NO: 1)/BDNF (SEQ ID NO: 2); AF086035 (SEQ ID NO: 3)/MAPK1IP1L (SEQ ID NO: 4); AK093366 (SEQ ID NO: 5)/AG2 (SEQ ID NO: 6); BC047792 (SEQ ID NO: 7)/PURB (SEQ ID NO: 8); AK096235 (SEQ ID NO: 9)/LCP1 (SEQ ID NO: 10); AL110130 (SEQ ID NO: 11)/SMEK2 (SEQ ID NO: 12); BC013641 (SEQ ID NO: 13)/ARC (SEQ ID NO: 14); hTF27297 (SEQ ID NO: 15)/CYR61 (SEQ ID NO: 16); RPPH1 (SEQ ID NO: 17)/NEAT1 (SEQ ID NO: 18); NEAT1 (SEQ ID NO: 18)/EGR3 (SEQ ID NO: 19); NEAT1 (SEQ ID NO: 18)/CR600638 (SEQ ID NO: 20); RPPH1 (SEQ ID NO: 17)/NEAT2 (MALAT1) (SEQ ID NO: 21); CR600638 (SEQ ID NO: 20)/FLT1 (SEQ ID NO: 22); BC012463 (SEQ ID NO: 23)/LRRN1 (SEQ ID NO: 24); CR615000 (SEQ ID NO: 25)/ERAP1 (SEQ ID NO: 26); BCO28229 (SEQ ID NO: 27)/BTG3 (SEQ ID NO: 28); BC078172 (SEQ ID NO: 29)/VIM (SEQ ID NO: 30); AF075087 (SEQ ID NO: 31)/S1PR1 (SEQ ID NO: 32); AK123944 (SEQ ID NO: 33)/TRIM47 (SEQ ID NO: 34); or AL110176 (SEQ ID NO: 35)/PLOD2 (SEQ ID NO: 36).

**[0020]** Another embodiment includes a method comprising targeting an IncRNA gene as a site of putative therapeutic intervention wherein the IncRNA gene is differentially expressed in at least one area of the brain having a different characteristic then a second area and wherein the IncRNA gene is RPPH1; NEAT1 or NEAT2.

**[0021]** Another embodiment includes a method comprising targeting a gene as a site of therapeutic intervention when the gene was identified by a method or microarray disclosed herein. In another embodiment, the targeting includes gene silencing or gene activating. In another embodiment, the targeting includes gene silencing through RNA interference (RNAi).

## BRIEF DESCRIPTION OF THE FIGURES

**[0022]** FIG. **1**. Reciprocal pattern of BDNF and BDNFOS gene expression in electrically active human neocortex. A. Summary of human epilepsy patients showing the ratios of electrical discharges, regions of neocortex sampled for each, and histopathology. All tissue sampled for gene expression changes had a normal histological structure, even in the presence of nearby structural abnormalities. B. Long-term brain surface recordings obtained prior to tissue resection were used to differentiate electrode locations with high- and low-spiking interictal activities for each patient. C. While both the activity-dependent immediate early gene EGR1 and BDNF are both constitutively upregulated in high-spiking cortex, BNDFOS was consistently downregulated in the same samples. The downregulation was significant ($p=0.016$, Wilcoxon's test; BDNFOS fold change $<-1.1$, 95% CI), Bars represent average values for all 7 patients shown with the shading of the circles corresponding to patients shown in A.

**[0023]** FIG. **2**. Downregulation of BDNFOS induces BDNF expression in Sy5Y cells. A. The BDNFOS/BDNF gene locus shows antisense overlap between BDNF and BDNFOS (UCSC Genome Browser). Three siRNAs were generated from a non-overlapping BDNFOS exon (s1, s2, s3). B. Downregulation of BDNFOS IncRNA using each of these siRNAs produced a corresponding increase in BDNF mRNA levels at 24 h and 48 h. Expression level changes are relative to a mock-electroporation negative control. Standard error bars displayed. No further BDNFOS knockdown or BDNF rescue persisted at the 48 h time point for s1 and s3 (data not shown).

**[0024]** FIG. **3**. Genome-wide analysis of human cortex reveals activity-dependent gene pairs and standalone IncRNAs. A. This experimental design of paired high- and low-spiking brain samples from the 7 patients shown in FIG. **1***a* was used both to interrogate coding and non-coding gene transcription as a function of brain activity. A dye-flip, quadruplicate microarray design was used with both a genome-wide coding array and a custom IncRNA array encompassing 5586 IncRNA genes with 7 probes/gene. Based on a rigorous statistical cutoff, a total of 4044 protein-coding and 1288 IncRNA genes were initially identified for these 7 patients (>1.4 fold and FDR<5% for each probe). LncRNA genes were further subdivided based on known cis-antisense partners of mRNAs, IncRNAs located <10 kb from any known gene, or standalone IncRNAs >10 kb from any known gene. Due to gene chains, some IncRNAs belonged simultaneously to the first two of these three categories. B. Pairs of differentially expressed mRNA and IncRNA genes that were either in an antisense overlap or <10 kb neighbors (which we define as "cis-encoded pairs") are shown together with IncRNAs that reside at other types of genomic loci. Many of the affected mRNAs have known functions in synaptic plasticity. The arrows to the left of the gene/pairs have been validated by qPCR. *BDNFOS/BDNF was discovered by targeted qPCR and did not meet statistical significance on the microarrays. Note that since the creation of this Figure, some "standalone" IncRNAs have been paired through additional analysis. To date these IncRNAs include RPPH1, NEAT1 and NEAT2 (MALAT1).

**[0025]** FIG. **4**. Parallel patterns of activity-dependent gene pairs. As a means to identify activity-dependent IncRNAs with potential roles in synaptic plasticity, the expression patterns of 13 known activity-dependent coding genes against the entire dataset of IncRNAs were probed for parallel patterns of expression. This figure shows significant relationships between these 13 genes and 26 IncRNAs identified using an R>0.90 cutoff. Each line represents a significant correlation and the proximity of the genes is directly proportional to this significance. The length of each line is inversely proportional to the correlation coefficient that is based on the average of correlations from probes above the 0.90 cutoff. The width of each line is directly proportional to the number of probes above the 0.90 cutoff. Coding genes are shown in dark grey while IncRNAs are in lighter grey. This figure was prepared using Cytoscape (http://www.cytoscape.org).

**[0026]** FIG. **5**. Repeated depolarization in vitro can replicate patterns of coding/non-coding gene transcription. A.

While transient CREB phosphorylation is induced with a single depolarization of Sy5Y cells with 100 mM KCl producing downregulation of BDNF and BDNFOS, B. repeated depolarization produces more sustained CREB activation and results in a reciprocal pattern of BDNF/BDNFOS transcription at 24 hours (shown by the brackets). For each of these, the middle panel shows a quantitation of triplicate Western blots as well as triplicate qPCR results for EGR1 to show activity dependent transcription, together with BDNF and BDNFOS at each time point. C. The expression of three cis-encoded lncRNA/mRNA pairs and one known functional lncRNA (NEAT1) was examined in the same SY5Y repeated depolarization time course as in (b) showing widely different patterns of expression of lncRNAs located both within and outside of genomic loci that harbor cis-encoded lncRNA/mRNA pairs (as used herein, "cis-encoded" includes situations where an lncRNA and an mRNA or an lncRNA and an lncRNA are expressed from the same or adjacent genomic loci. Adjacent genomic loci include those with end nucleotides within 10 nucleotides of the other).

**[0027]** FIG. **6**. Genomic complexity of the human AK093366/AG2 lncRNA/mRNA cis-antisense pair which is co-differentially expressed in human neocortical epilepsy.

**[0028]** FIG. **7**. Taqman qRTPCR results closely parallel microarray results for lncRNA and mRNA differential expression at lncRNA/mRNA cis-pairs across the within-patient sample pairs of high- and low-activity neocortical regions. MALAT-1 is not shown because of the discrepancy between its microarray probeset coverage and its Taqman amplicons coverage.

## DETAILED DESCRIPTION

**[0029]** The present disclosure provides activity-dependent gene pairs as therapeutic targets and methods and devices to identify the same. The present disclosure also describes the first genome-wide analysis of human brain long non-coding RNA (lncRNA)-based gene pairs as a function of coding mRNA or other lncRNA pairings and electrical brain activity. Many of the coding mRNAs identified in this way are known to modulate activity-dependent gene expression in the human brain, suggesting that these particular lncRNA/mRNA pairs form regulatory networks related to human brain plasticity. LncRNA/lncRNA pairs showing differential expression between areas of high and low brain activity were also observed. The described lncRNA/mRNA and lncRNA/lncRNA pairs provide targets for rational therapeutic development in a number of disorders and diseases associated with differential or perturbed electrical brain activity including epilepsy. As used herein, "differential expression", "significantly differentially expressed" and similar terms mean that expression of a gene is significantly different based on a statistical power analysis, the results of which can be validated by qPCR at a 95% confidence interval. "Expression" as used herein includes (i) transcription of a gene encoding lncRNA and (ii) transcription of a gene encoding protein-encoding RNA and/or translation of the protein-coding RNA. "High" and "low" electrical brain activity, interictal and/or ictal spiking as used herein are relative terms to be compared between brain areas in a given patient during a given procedure.

**[0030]** The abundance of non-protein-coding transcriptional units rivals the numbers of known mRNAs. LncRNA genes can be defined by four criteria: encoding transcripts that lack any open reading frames (ORFs) greater than 100 amino acids or possessing protein database homologies; being within the known range of lengths of mammalian mRNAs (from ~300 nt to >20,000 nt in length); support by transcript-to-genome alignments from cDNA data; and absence of matches to any known non-coding-RNA classes. Further information regarding lncRNAs may be found by consulting the GENCODE Project which annotates evidence-based gene features of the human genome including protein-coding loci with alternatively-transcribed variants, non-coding loci with transcript evidence and pseudogenes. GENCODE has assigned the lncRNA biotype to certain genes and transcripts, signifying that those genes and transcripts encode and represent, respectively, lncRNAs based on the best available manual-annotation and experimental data. Any gene or transcript given an lncRNA biotype by GENCODE is an lncRNA as the term is used herein regardless of whether the particular gene or transcript meets the four criteria provided above. The GENCODE gene sets are used by the ENCODE consortium.

**[0031]** Numerous lncRNAs are transcribed in the vicinity of known mRNAs, and regulate those known genes through epigenetic mechanisms. Functionally, lncRNAs can have regulatory effects on coding mRNAs through a number of mechanisms that include cis-antisense lncRNA transcripts that repress their sense-strand protein-coding partners. LncRNAs can also enhance expression of differentially expressed gene pair partners. LncRNAs encoded in an antisense orientation to, and overlapping with, known mRNAs are particularly abundant. The vast majority of these lncRNAs remain devoid of known functions.

**[0032]** The human brain is composed of a diverse set of cell types connected through complex synaptic arrangements. The degree of synaptic activity in the brain can be translated into functional and structural changes through activity-dependent changes in gene expression. Although these changes can be effected through direct activation of synaptic genes, they can also be achieved through the release of neurotrophic factors such as brain-derived neurotrophic factor (BDNF) that have direct effects on synaptic architecture and indirect effects by producing changes in gene expression. BDNF, a member of the nerve growth factor family, regulates the survival and differentiation of neuronal populations, axonal growth and pathfinding, dendritic growth and morphology and has been linked to many human brain disorders. BDNF mRNA and protein are upregulated by seizure activity in animal models of epilepsy as well as in human brain tissues that display increased epileptic activities. The genomic locus encoding BDNF is structurally complex and also encodes BDNFOS, a primate-specific lncRNA that is antisense to the coding BDNF gene. BDNF and BDNFOS form double-stranded duplexes, suggesting a potential for BDNFOS to post-transcriptionally regulate BDNF.

**[0033]** BDNF binding to its receptors results in a diverse array of downstream signaling pathways including the activation of cyclic adenosine monophosphate response element binding protein (CREB), that in turn can also regulate BDNF by binding to a cognate site within the BDNF gene. Activation of CREB by phosphorylation at Serine 106 as a result of neuronal activity leads to changes in gene expression that cause reinforcement and stabilization of more active neuronal circuits. Downstream from phosphorylated CREB (pCREB), immediate early genes (IEGs) have been shown to mediate long-lasting changes in neuronal structure and excitability. Upstream of CREB activation, several known signaling path-

ways are rapidly activated in response to neuronal activity, including CaMKinase IV, protein kinase A, and MAPK. A pattern of transcriptional activation in human brain regions where seizures start strongly implicates sustained MAPK/CREB activation and downstream coding gene activations that could underlie layer specific changes in synaptic architecture that makes these regions prone to seizures.

**[0034]** Given that human lncRNA genes tend to be less well-conserved than mRNAs, and have unique transcripts not found in other species, a uniquely human system to examine activity-dependent gene expression for both coding and non-coding RNAs using a pairwise comparison of human cortical regions displaying variable degrees of epileptic activities was sought. These brain regions were removed as part of surgical treatment for intractable seizures. It was shown that regions of human neocortex that display increased activity and BDNF expression have reduced BDNFOS expression and that BDNFOS directly down-regulates BDNF in vitro. A custom microarray platform to perform a transcriptome-wide analysis of other regulatory lncRNAs was also developed and matched to differentially expressed mRNAs or other lncRNAs to develop a genome-wide list of lncRNA/mRNA gene pairs and/or lncRNA/lncRNA gene pairs. Many of the coding mRNAs identified in this way are known to modulate activity-dependent gene expression in the human brain, suggesting that these lncRNA/mRNA pairs form a newly revealed regulatory network of human brain plasticity. Identified lncRNA/lncRNA pairs also provide targets for potential therapeutic intervention.

**[0035]** Genome-Wide Integration of Activity-Dependent Gene Pairs as a Function of Human Brain Activity.

**[0036]** The present disclosure describes a stringently filtered catalog of human lncRNAs and the genomic positional relationships between these lncRNAs and mRNAs and/or other lncRNAs, providing insights into lncRNA functions (see Jia et al. 2010, incorporated by reference for its teachings regarding the same). Despite their prominence in the transcriptome, most lncRNAs remain poorly understood. The present disclosure describes development of a custom lncRNA microarray to provide the first genome-wide analysis of human brain lncRNA-based gene pairs as a function of electrical brain activity. Several identified co-expressed lncRNA/mRNA gene pairs have important roles in activity-dependent synaptic plasticity either directly, such as BDNF and others involved in the MAPK/CREB signaling, or indirectly through the expression of regulatory lncRNAs such as MALAT-1 which are members of trans-encoded lncRNA/lncRNA gene pairs (e.g. RPPH1/MALAT-1, whereas MALAT-1 is an RNA-processing target of RPPH1).

**[0037]** The described genomewide lncRNA expression survey of electrically active human neocortex uncovered hundreds of lncRNAs differentially expressed between more and less electrophysiologically active areas of the human neocortex. Of these lncRNAs, 26 were initially identified as expressed directly in proportion to known activity-dependent genes (FIG. **4**). These lncRNAs represent therapeutic targets for human brain diseases, such as epilepsy. Without being bound by theory, the co-expression clustering topology (FIG. **4**) suggests a network where mRNAs and lncRNAs are linked by previously uncharacterized lncRNA nodes (such as BC009864) as hubs with spoke edges extending simultaneously to multiple mRNAs and other lncRNAs. Eight lncRNA/mRNA cis-antisense and neighbor-gene pairs characterized by coordinated differential expression of both genes in each lncRNA/mRNA pair were also initially observed, suggesting lncRNA-mediated regulation of protein-coding gene expression in the human brain. These pairs also suggest that some mRNAs function at the RNA level to regulate lncRNA expression or in bidirectionally regulated feedback loops in cis. Other lncRNAs such as NEAT1 were detected only by the trans-regulation analysis, which targeted lncRNAs whose expression was highly correlated with mRNAs or other lncRNAs regardless of the genomic mapping location of those coding genes. The trans-regulation analysis implies NEAT1, the lncRNA from nuclear paraspeckles that is encoded near the NEAT2 locus, in regulatory interactions with activity-dependent genes in the brain. Three lines of evidence for activity-dependent NEAT1 function in the neocortex include (1) detection of NEAT1 as a differentially expressed lncRNA on the custom microarray analysis of human brain samples, (2) demonstration of activity-dependent NEAT1 expression in depolarized human SY5Y cell culture, and (3) the assignment of NEAT1 as a central node to a trans-encoded co-expression cluster of specific coding and non-coding RNAs (FIG. **4**). The described cis-regulation and trans-regulation analyses uncovered different, nonredundant sets of lncRNAs, suggesting that specific lncRNAs are involved in both types of regulation, which for any given lncRNA may be mutually exclusive. These results represent the first functional evidence for a remarkably diverse pattern of lncRNA expression and function in the human brain.

**[0038]** Functionality of Activity-Dependent Gene Pairs in the Human Brain.

**[0039]** The described microarray results and qPCR analysis of both the epilepsy patient samples and the recurrent-depolarization Sh-sy5y cell culture timecourse, jointly represent the first demonstration that known lncRNAs are activity-dependent both in vivo and in cell culture. The complex, but similar, pattern of lncRNA/mRNA or lncRNA/lncRNA expression in activated human brain and in a chronically depolarized human neuronal cell line enables the temporal characterization of these regulatory gene pairs and provides a new system in which to assess these complex, primate-specific transcriptional gene pairs.

**[0040]** The present disclosure describes upregulation of three nuclear RNAs, RNase P (RPPH1), NEAT1, and MALAT-1, in high-activity areas of the neocortex. The catalytic-RNA component of RPPH1 is essential for the 3' end cleavage of both NEAT1 and NEAT2/MALAT-1. Therefore, these three lncRNAs may comprise an lncRNA-mediated lncRNA maturation network in highly active brain regions. The function of this induced network is predicted to modulate the expression of synaptic genes, such as those whose mRNA levels are regulated by NEAT2/MALAT-1. This RNA-mediated regulatory network is also predicted to function either independently from, or synergistically with, the MAPK/pCREB pathway to regulate activity-dependent gene expression.

**[0041]** Ribonucleoprotein complexes that enable lncRNA function and complexes which facilitate lncRNA-mediated regulation of mRNAs in sense-antisense pairs can be identified by affinity columns and mass spectrometric analysis. This identification will allow therapeutic targeting of lncRNA/mRNA and/or lncRNA/lncRNA gene pairs.

**[0042]** The genomic complexity of the human AK093366/AG2 lncRNA/mRNA cis-antisense pair (FIG. **6**) is reminiscent of that observed in the BDNFOS/BDNF lncRNA/mRNA cis-antisense pair. Both the BDNFOS lncRNA gene

and the AK093366 lncRNA gene contain primate-specific splice sites. The splice donor of AK093366's sole intron is primate-specific because it is harbored within an AluJb repeat. Alu repeats are the best-known class of primate-specific interspersed repeats and therefore key gene structure elements, including splice sites, contained within Alu repeats provide direct evidence that the corresponding gene structures either arose or were modified after the mammalian radiation, specifically in the primate lineage. Notably, EST-supported cis-antisense lncRNA transcription of the Alu-containing AK093366 transcriptional unit extends substantially beyond the UCSC C110RF96 (AG2) gene model, and well into the C110RF96 ORF. This underscores the utility of EST data, much of which remains unincorporated into reference gene models and annotations, in delineating the boundaries of lncRNA genes, including those involved in putative regulatory relationships with protein-coding counterparts. While Alu-containing lncRNAs have been implicated in the in-trans post-transcriptional regulation of gene expression via effecting mRNA decay, the currently described analysis suggests distinct, cis-regulatory roles in overlapping-gene regulation for certain Alu-containing lncRNAs, specifically AK093366. Two co-differentially-expressed lncRNA/mRNA cis-antisense pairs in the human neocortex, BDNFOS/BDNF and AK093366/AG2, thus feature primate-specific sequence at lncRNA gene splice junctions. mRNAs BDNF and AG2 are overlapped by endogenous antisense lncRNAs containing exonic Alu repeats, and these gene pairs are cod ifferentially-expressed in active areas of the human epileptic neocortex.

**[0043]** BDNFOS-mediated regulation of BDNF provides evidence that primate-specific regulation of conserved mRNAs by cis-antisense lncRNAs takes place in epilepsy, a human brain disorder. The co-expression of mRNAs and non-conserved lncRNAs at loci other than BDNF, including AG2, demonstrates that primate-specific regulation of conserved genes by nonconserved lncRNAs in the human brain is not unique to the BDNF locus.

**[0044]** A Primate-Specific lncRNA Regulatory Mechanism for BDNF.

**[0045]** A striking feature of the BDNF/BDNFOS locus is the complexity of its genomic landscape, which is highly representative of the genomic properties observed at lncRNA-encoding loci throughout mammalian genomes. Human BDNFOS is part of a three-gene genomic positional chain: it shares a putative bidirectional promoter with the LIN7C gene at its 5' end, while sharing its exonic cis-antisense overlap with BDNF exonic sequences at the 3' end. In addition, BDNFOS may have emerged in recent mammalian evolution, after the primate-rodent divergence. A possible recent origin for this lncRNA gene is supported by two lines of evidence. First, several splice sites of BDNFOS are poorly conserved outside of primates, suggesting that the genomic structure of BDNFOS either arose or was modified specifically in primate evolution. This is consistent with the recent finding that lncRNA genes may comprise a majority of primate-specific genes. Second, there is no evidence for a BDNFOS-like gene between mouse Lin7c and mouse BDNF in mouse cDNA and EST sequence data (UCSC transcript-to-genome alignments). This genomic and evolutionary complexity of the BDNF/BDNFOS locus, implies that functional lncRNAs in the human brain may be characterized by inter-species non-conservation or high divergence of their gene structures. This is of particular interest because of the persistent inverse co-differential-expression of the BDNF/BDNFOS gene pair as a function of human brain activity shown here together with the observed increase in BDNF mRNA levels following knock down of BDNFOS. The present disclosure therefore provides a uniquely human view of activity-dependent gene pairs in the brain whose endogenous components cannot be modeled in rodents or other non-primate species. A set of differentially expressed miRNAs, including miR-30a-5p, act as post-transcriptional inhibitors of BDNF in prefrontal cortex. The demonstration of the primate-specific lncRNA BDNFOS as a post-transcriptional inhibitor of BDNF complements this miRNA work, suggesting that BDNF is targeted by multiple RNA-mediated regulatory mechanisms involving short and long, ancient as well as evolutionarily young ncRNAs.

**[0046]** The methods and devices described herein identified the following lncRNA/mRNA or lncRNA/lncRNA gene pairs as activity-dependent targets for therapeutic intervention (lncRNA/mRNA or lncRNA/lncRNA): BDNFOS (SEQ ID NO: 1)/BDNF (SEQ ID NO: 2); AF086035 (SEQ ID NO: 3)/MAPK1IP1L (SEQ ID NO: 4); AK093366 (SEQ ID NO: 5)/AG2 (SEQ ID NO: 6); BC047792 (SEQ ID NO: 7)/PURB (SEQ ID NO: 8); AK096235 (SEQ ID NO: 9)/LCP1 (SEQ ID NO: 10); AL110130 (SEQ ID NO: 11)/SMEK2 (SEQ ID NO: 12); BC013641 (SEQ ID NO: 13)/ARC (SEQ ID NO: 14); hTF27297 (SEQ ID NO: 15)/CYR61 (SEQ ID NO: 16); RPPH1 (SEQ ID NO: 17)/NEAT1 (SEQ ID NO: 18); NEAT1 (SEQ ID NO: 18)/EGR3 (SEQ ID NO: 19); NEAT1 (SEQ ID NO: 18)/CR600638 (SEQ ID NO: 20); RPPH1 (SEQ ID NO: 17)/NEAT2 (MALAT1) (SEQ ID NO: 21); CR600638 (SEQ ID NO: 20)/FLT1 (SEQ ID NO: 22); BC012463 (SEQ ID NO: 23)/LRRN1 (SEQ ID NO: 24); CR615000 (SEQ ID NO: 25)/ERAP1 (SEQ ID NO: 26); BCO28229 (SEQ ID NO: 27)/BTG3 (SEQ ID NO: 28); BC078172 (SEQ ID NO: 29)/VIM (SEQ ID NO: 30); AF075087 (SEQ ID NO: 31)/S1PR1 (SEQ ID NO: 32); AK123944 (SEQ ID NO: 33)/TRIM47 (SEQ ID NO: 34); and AL110176 (SEQ ID NO: 35)/PLOD2 (SEQ ID NO: 36).

**[0047]** In a preferred embodiment, the first lncRNA of a pair is targeted to effect therapeutic intervention. In alternative embodiments, the mRNA and/or the second lncRNA of a pair can be targeted. In further embodiments, the first lncRNA and mRNA or second lncRNA of a pair can both be targeted.

**[0048]** The genes of the above pairs to target include the sequences above as well the reverse complements thereof and allelic variants. Allelic variants include slightly different sequences that originate from the same chromosomal position or the same position on an allelic chromosome. Therapeutic targets as disclosed herein also include sequences with at least 90% sequence identity; at least 91% sequence identity; at least 92% sequence identity; at least 93% sequence identity, at least 94% sequence identity, at least 95% sequence identity, at least 96% sequence identity, at least 97% sequence identity, at least 98% sequence identity or at least 99% sequence identity to SEQ ID NO. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35 and/or 36. Percentage of sequence identity is determined by comparing two optimally aligned sequences (e.g., nucleic acid sequences) over a comparison window, wherein the portion of the sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the

number of positions at which the identical nucleotide or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the comparison window, and multiplying the result by 100 to yield the percentage of sequence identity. A sequence that is identical at every position in comparison to a reference sequence is said to be 100% identical to the reference sequence, and vice-versa. Therapeutic targets as disclosed herein also include those sequences that hybridize to one or more of SEQ ID NO. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35 and/or 36 under high stringency conditions. As used herein, high stringency conditions include hybridization in 5×SSC buffer at 65° C. for 16 hours; wash twice in 2×SSC buffer at room temperature for 15 minutes each; and wash twice in 0.5×SSC buffer at 65° C. for 20 minutes each.

**[0049]** As will be understood by one of ordinary skill in the art, the therapeutic targets disclosed herein can be targeted through any mechanism capable of increasing or decreasing their function, as appropriate. Many of these strategies will be based on complementary binding properties, but the present disclosure is not so limited and can include any form of effective intervention, however formed and delivered. The currently-disclosed targets can also be targeted by inhibiting or increasing the activity of upstream or downstream molecules, including those described herein in relation to particular targets including CREB, CaM Kinase IV, protein kinase A and MAPK. While not limiting the foregoing inclusive statements in any manner, the following description provides several appropriate targeting strategies.

**[0050]** One targeting approach can include gene silencing. This approach refers to the reduction in transcription, translation, expression or activity of a nucleic acid, as measured by transcription level, mRNA or lncRNA level, enzymatic activity, methylation state, chromatin state or configuration, translational level, or other measure of its activity or state in a cell or biological system. Such activities or states can be assayed directly or indirectly. Gene silencing also includes the reduction or amelioration of activity associated with a nucleic acid sequence, such as its ability to function as a regulatory sequence, its ability to be transcribed, its ability to be translated and result in expression of a protein, regardless of the mechanism whereby such silencing occurs.

**[0051]** One method of gene silencing includes RNA interference (RNAi). RNAi refers to the process by which a polynucleotide or double stranded polynucleotide comprising at least one ribonucleotide unit exerts an effect on a biological process through disruption of gene expression. The process includes but is not limited to gene silencing by degrading mRNA, interactions with tRNA, rRNA, hnRNA, cDNA and genomic DNA, as well as methylation of DNA and ancillary proteins.

**[0052]** Another targeting approach can include gene activating. This approach refers to an increase in transcription, translation, expression or activity of a nucleic acid, as measured by transcription level, mRNA or lncRNA level, enzymatic activity, methylation state, chromatin state or configuration, translational level, or other measure of its activity or state in a cell or biological system. Such activities or states can be assayed directly or indirectly. Furthermore, gene activating includes the increase of activity associated with a nucleic acid sequence, such as its ability to function as a regulatory sequence, its ability to be transcribed, its ability to be translated and result in expression of a protein, regardless of the mechanism whereby such activation occurs.

**[0053]** Any type of nucleic acid capable of achieving gene silencing or gene activation can be used whether such nucleic acids are endogenously, exogenously and/or recombinantly-derived. Exemplary types of nucleic acid molecules that can be used in targeting strategies disclosed herein include:

**[0054]** Short interfering RNA (siRNA)—a double stranded nucleic acid that is capable of performing RNAi and that is between 18 and 30 base pairs in length. siRNAs can be duplexes, and can also comprise short hairpin RNAs, RNAs with loops as long as, for example, 4 to 23 or more nucleotides, RNAs with stem loop bulges, micro-RNAs, and short temporal RNAs. RNAs having loops or hairpin loops can include structures where the loops are connected to the stem by linkers such as flexible linkers. Flexible linkers can be comprised of a wide variety of chemical structures, as long as they are of sufficient length and materials to enable effective intramolecular hybridization of the stem elements. Typically, the length to be spanned is at least about 10-24 atoms. siRNAs can be endogenous or exogenous, although in practice, therapeutic siRNA will be exogenous.

**[0055]** MicroRNA (miRNA)—18-25 nucleotide non-coding RNAs derived from endogenous genes. MiRNAs assemble in complexes and recognize their targets by antisense complementarity. If the miRNA matches its target with 100% sequence identity, the target RNA is cleaved, and the miRNA acts like a siRNA. If the sequence identity is less than 100%, the translation of target RNA is blocked.

**[0056]** Small nucleolar RNAs (snoRNAs)—small RNA molecules that guide chemical modifications (methylation or pseudouridylation) of ribosomal RNAs (rRNAs) and other RNA genes (tRNAs and other small nuclear RNAs (snRNAs)).

**[0057]** As provided, a variety of types of nucleic acid molecules can be used to increase or decrease the function of the therapeutic targets identified herein. The nucleic acids can include either or both naturally occurring and modified nucleotides linked together by naturally occurring and/or non-naturally occurring nucleotide linkages. Nucleic acid molecules may be modified chemically or biochemically, or may contain non-natural or derivatized nucleotide bases. Such modifications include, for example, labels, methylation, substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications (e.g., uncharged linkages: for example, methyl phosphonates, phosphotriesters, phosphoramidates, carbamates, etc.; charged linkages: for example, phosphorothioates, phosphorodithioates, etc.; pendent moieties: for example, peptides; intercalators: for example, acridine, psoralen, etc.; chelators; alkylators; and modified linkages: for example, alpha anomeric nucleic acids, etc.). Additionally, the nucleic acid molecules may be modified at either the 3' and/or 5' end by any type of modification known in the art. For example, either or both ends may be capped with a protecting group, attached to a flexible linking group, attached to a reactive group to aid in attachment to a substrate, etc. The term "nucleic acid molecule" also includes any topological conformation, including single-stranded, double-stranded, partially duplexed, triplexed, hairpinned, circular, and padlocked conformations.

**[0058]** Conjugates may also be used to target the gene pairs described herein. As used herein, conjugates refer to molecules with at least two discrete components. In one embodiment, one component alters the physical properties of another

component. The altered physical property can be, without limitation, shelf-life stability, in vivo half-life or cellular uptake. In a particular embodiment, the two components are covalently or non-covalently associated. In another embodiment, one component is a nucleic acid molecule and the second component is a non-nucleotide region such as, without limitation, polyethylene glycol, one or more fatty acid chains, one or more sugar residues, etc.

**[0059]** Pharmaceutically acceptable excipients such as vehicles, adjuvants, carriers and/or diluents can be used as appropriate Other components such as pH adjusting and buffering agents, tonicity adjusting agents, stabilizers, wetting agents and the like can also be used.

## Examples

### Experimental Rationales, Materials and Methods

**[0060]** Human Brain Tissue:

**[0061]** Patients who fail to respond to medical management of their seizures can greatly benefit from a 2-stage surgical procedure where long-term in vivo brain surface recordings are used to identify and remove epileptic brain regions. While removing seizure onset regions is key to a good outcome for improved seizure control, seizures from these brain regions are relatively infrequent compared to the small, but extremely frequent 'interictal' epileptic discharges that can occur almost constantly between seizures in some brain regions. In fact, several of the genes induced at seizure onset zones correlate precisely with interictal spiking rather than seizure frequency, suggesting that interictal spiking may be the driving force behind this altered expression pattern. Consistently, an animal model of interictal spiking without seizures was sufficient to produce neuronal layer-specific changes in these genes. The studies of the present disclosure have focused on brain regions with different levels of brain activity as measured by interictal spiking to identify the relationships between coding and non-coding or non-coding and non-coding transcripts in the human brain.

**[0062]** Informed consent was obtained from seven patients who underwent surgery for medically intractable epilepsy (FIG. **1***a*). Extreme care was taken to ensure that the described study did not influence surgical decision-making. All patients underwent presurgical evaluation and identification of epileptic and control regions as previously described (Beaumont in revision 2011; Rakhade et al. 2005 both incorporated by reference herein for their teachings regarding the same). To localize epileptic brain regions that displayed both clinical seizures and interictal epileptiform discharges (spikes), a two-stage surgical approach using subdural electrodes with continuous brain surface recordings (electrocorticography) and video monitoring was undertaken for two to five days. For these studies, paired tissue samples from neocortex within each patient displaying high and low amount of interictal (between seizures) spiking as determined by quantitative intracranial electrode recordings were used to compare differential gene expression as a function of brain activity (Loeb 2010, incorporated by reference herein for its teachings regarding the same; see also Barkmeier et al., 2012 incorporated by reference herein for its teachings regarding inter-reviewer variability of spike diction on intracranial EEG addressed by an automated multi-channel algorithm). To avoid introducing additional variables into the analysis, each block of tissue was examined histopathologically, and demonstrated a normal 6-layered neocortical structure without lesions. The paired analysis within each patient is critical to isolate the variable under study, which is the degree of activity. Total RNA was prepared using a modification of the protocol described previously (Beaumont in revision 2011, incorporated by reference herein for its teachings regarding the same). The difference was that only gray matter was used by pooling 2-3 nearby strips of gray matter that extended from the pial surface to the white matter from each block of tissue corresponding to a given electrode location. This pooling method helps correct for differences in dissections that could lead to over- or under-representation of specific cortical layers.

**[0063]** Cell Cultures, Transfections, and Depolarizations:

**[0064]** The SH-SY5Y cell line (ATCC) was maintained in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% FBS and used for experiments. Cells between 17 and 25 passages were transfected with BDNFOS-targeting and BC013641-targeting siRNAs by electroporation according to manufacturer's instructions at approximately 80% confluence (Neon electroporation system, Invitrogen). The electroporation conditions used for SH-SY5Y cell transfection were 1200V, 20 Pulse Width, 2 Pulse numbers, which were optimized using a condition matrix, a control siRNA, and fluorescent reporters (data not shown). Single and multiple depolarizations of cells were induced by adding 100 mM KCl (final concentration) to the medium at different time points as indicated in the figure legend.

**[0065]** gPCR, siRNAs, and Primers:

**[0066]** Total RNA from cultured SH-SY5Y cells was isolated with RNA easy Mini kit according to manufacturer's instructions (QIAgen). The first-strand cDNA was prepared using SuperScript First-Strand cDNA kit (Invitrogen), mRNA and IncRNA expression levels were determined by Taqman quantitative real-time PCR (Taqman qRTPCR). BDNFOS siRNAs named 51, S2, S3, and S4 were custom-designed and synthesized by Invitrogen. The BDNFOS Taqman primer/probe combos were custom-designed by uploading FASTA-format sequences of preferred amplicons regions the ABI Taqman custom design website, and purchased from ABI/Life Technologies. This vendor does not release the actual primer and probe sequences of custom-designed amplicons to the users.

**[0067]** Western Blot Analysis:

**[0068]** Cell lysates were prepared with SDS sample buffer (Sigma) and subjected to Western blotting to measure CREB phosphorylation as described (BEAUMONT in revision 2011, incorporated by reference for its teachings regarding the same). Briefly, proteins separated on 4-20% gradient sodium dodecyl sulfate-polyacrylamide gel were electrically transferred onto nitrocellulose membrane. After blocking with 5% (v/v) skim milk in TBS containing 0.05% Tween-20 (TBST) for 1 hour at room temperature (RT), the membrane was incubated with rabbit polyclonal antibody against p-CREB (Cell Signaling) at dilution of 1:1000 for 1 hour at RT and then with specific secondary antibody coupled with HRP (1:5000) for 1 hour at RT. p-CREB was visualized with ECL detection system (Pierce). The membrane was then stripped and re-probed with CREB antibody (Cell Signaling) at (1:1000) to measure total CREB.

**[0069]** Microarrays:

**[0070]** Seven 60-mer probes per gene, unambiguously mapping by BLAT (KENT 2002, incorporated by reference herein for its regarding the same) to a single genomic loca-

tion, and free of interspersed and simple repeats, were designed using the Agilent Technologies OpenGenomics eArray interface for 5586 of the 6736 lncRNA genes from Jia et al. (Jia et al. 2010, incorporated by reference herein for its teachings regarding the same). The remaining lncRNA genes were excluded because of eArray failure to yield 7 probes per gene, or because the eArray-designed probes failed a subsequent check for genomic uniqueness and absence of repeats. As a positive control, seven probes each for 111 of the 137 previously determined protein-coding epileptic genes (BEAUMONT in revision 2011) and for 6 housekeeping control genes were included. The eArray Fill Array feature was used to randomly select control protein-coding gene probes to fill all features that would have otherwise remained vacant (<2% of total features on a 44,000-feature, i.e. "44 k," array cell). The entire probeset was printed in quadruplicate on each slide using the Agilent 4×44 k high-density oligonucleotide microarray platform.

**[0071]** A dye-flip quadruplicate two-color microarray experiment was performed on each within-patient pair of high-spiking and low-spiking surgically resected samples on both the Agilent human genome-wide array (G4112A) and our custom lncRNA array as described, but using a different labeling method (BEAUMONT in revision 2011). The Epicentre protocol was used to generate aminoallyl-aRNA for subsequent amplification and labeling with either Cy or Alexa dyes. The custom lncRNA arrays were labeled with Alexa dyes (Alexa-647 and Alexa-555, Invitrogen) within the dye-flip design, as described by the manufacturer (SuperScript Indirect RNA Amplification System, Invitrogen) (Holloway et al. 2008, incorporated by reference herein for its teachings regarding the same). For every patient, each of the quadruplicates was hybridized on four separate slides. Four slides of 4×44 k Agilent arrays (4 arrays, each composed of the same set of approximately 44,000 probes) were used to screen seven patients. All slides were scanned as described previously (BEAUMONT in revision 2011).

**[0072]** Because the described lncRNA custom microarray platform was new, qPCR was also used to validate a representative subset of differentially-expressed lncRNAs from each of the 3 categories that are indicated by the arrows in FIG. **3***b*. Which specific probes were responsible for the differential expression of each coding and non-coding gene observed across all seven patients was considered, and probes to target only the region of each transcript that was overlapped by the differentially expressed probes were used. Positive correlation coefficients were seen in all cases, ranging from 0.61 to 0.96 between the array and qPCR results within each patient; all protein-coding gene differential expression results were from the G4112A or F catalog protein-coding microarray and all lncRNA results were from the described lncRNA custom microarray (FIG. **6**).

**[0073]** Microarray Statistical Methods:

**[0074]** In order to identify those differentially expressed lncRNAs that may be directly regulating their overlapping or neighboring mRNAs, the described custom lncRNA expression microarray data was integrated with conventional mRNA expression microarray data for the in vivo high/low-activity cortical sample pairs from all 7 patients analyzed with both array types (FIG. **3***a*). For each epilepsy patient, there was a within-patient sample pair of a high-spiking and a low-spiking region. This within-patient sample pair was analyzed, using the same dye-flip quadruplicate strategy, for both the catalog coding (G4112A) and the custom lncRNA microarray. Differentially expressed genes were identified from both microarray platforms but using the same strategy. Consistency between arrays was first examined by correlating the fold-change of all protein-coding control genes common to both arrays, which was possible because the 111 'epileptic transcriptome' genes from prior protein-coding array work (BEAUMONT in revision 2011, incorporated by reference for its teachings regarding the same) were used as controls on the lncRNA array. The average value of the seven probes corresponding to each control gene on the lncRNA custom array was used. For 140 catalog (Agilent G4112A) coding-array probes corresponding to these 111 genes, Pearson's correlation coefficient was 0.90, attesting to very high reproducibility between the coding array and the non-coding custom array.

**[0075]** Scanned microarray images from coding and non-coding microarrays were analyzed by the software Agilent Feature Extraction (Agilent, V10.3.1)' with the default protocol 'GE2_107_Sep09'. A fluorescent correction factor was determined using both Q-RTPCR and Agilent Spike-IN probes. This correction factor was then applied on the fluorescence intensity (fluorescence at exponent 1.125) and improved the fold change prediction. The fluorescence distribution inside each repetition of the microarray experiments was normalized by 'R' V2.11 (R DEVELOPMENT CORE TEAM 2010) using the library limma' (Smyth and Speed 2003, incorporated by reference herein for its teachings regarding the same) in a two step process: (i) normalization of the intensity of fluorescence between dyes using a Loess correction (iterations=50, span=0.05), and (ii) scaling of the fluorescence intensity on the same range across all the arrays for each dye independently using quantil normalization. The quality of the normalization process of the microarray fluorescence was validated using MA plotdensity and distribution analysis. Normality was asserted using Anderson-Darling test from the library 'Nortest' (Gross 2006, incorporated by reference herein for its teachings regarding the same). For each array, the background level was globally computed using the median of the fluorescence intensity of the negative control probes and subtracted to the signal of each probe.

**[0076]** Once normalized, the microarrays were further analyzed using standard statistical methods (Kerr and Churchill 2001 b; Wolfinger et al. 2001, both incorporated by reference herein for their teachings regarding the same). The differentially expressed genes between high and low spiking areas of the brain were determined using a two-step mixed model analyze of variance (Jin et al. 2001, incorporated by reference herein for its teachings regarding the same) with the library LME4 (Bates 2009, incorporated by reference herein for its teachings regarding the same). This mixed model approach has been used to compute the fitted effect and the random effects simultaneously (Littell 1996, incorporated by reference herein for its teachings regarding the same). To improve the sensibility of the analysis (Jin et al. 2001; Kerr et al. 2000, both incorporated by reference herein for their teachings regarding the same), computation did not use the ratio but instead used dye fluorescence intensity indexed by the type of RNA (Tanaka et al. 2000, incorporated by reference herein for its teachings regarding the same) (RNA from high spiking area or RNA from low spiking area). The FDR and corrected p-Value for each gene was computed with 'R' using the library 'fdrtool' (Strimmer, 2009, incorporated by reference herein for its teachings regarding the same). Differentially expressed genes were detected using fold change and signifi-

cance simultaneously (Tanaka et al. 2000, incorporated by reference for its teachings regarding the same) and were determined as significantly differentially expressed if their fold change, for at least one probe per gene, was above or equal to 1.4 and if their FDR was equal or lower than 5% in a groupwise analysis with all 7 patients required. These thresholds were selected based on a power analysis using this dye-flip quadruplicate design (LOEB 2009, incorporated by reference for its teachings regarding the same).

**[0077]** To integrate the coding and non-coding transcriptomes of the human neocortex, differentially expressed mRNAs encoded by genomic loci overlapping, or adjacent to, the loci which also encoded differentially expressed lncRNA genes were identified as outlined in FIG. **3***a*. Specifically, cis-encoded gene pairs in which both a protein-coding gene and a non-coding (lncRNA) gene were expressed from the same locus were identified. These pairs are referred to as lncRNA/mRNA gene pairs. The pairs were then separated into two categories 'antisense' and 'neighbor,' both of which carry the potential for mRNA regulation by a paired lncRNA (Jia et al. 2010; Lipovich et al. 2010, both incorporated by reference for their teachings regarding the same). A cis-antisense gene pair was defined as two genes transcribed from the opposite strands of the same locus in a configuration such that at least some sequence in at least one exon overlaps one exon of the other gene. A neighbor-gene pair was defined as any gene pair such that the nearest boundaries of two nearby, but non-overlapping, genes are less than 10 kb away from one another.

**[0078]** In addition to a number of custom approaches to identify cis-acting coding lncRNA pairs, trans-acting lncRNAs were identified as significant and activity-dependent by their tight correlation (Pearson's correlation coefficient minimum of 0.9) to a well-known group of activity-dependent mRNAs (Beaumont in revision 2011; Rakhade et al. 2007, both incorporated by reference for their teachings regarding the same), which themselves had been co-expressed with a Pearson's correlation coefficient of 0.95. These results were displayed graphically using Cytoscape (Smoot et al. 2011, incorporated by reference for its teachings regarding the same). To include a trans-acting lncRNA in this group, at least one probe (of the seven available probes) representing the lncRNA gene had to meet this statistical requirement.

**[0079]** Co-differential expression was defined as a differential expression profile of two genes such that the differential expression of one gene was either inversely or directly, correlated with the differential expression of the other gene across multiple sample pairs, each of which originated from a different patient and all of which were statistically significant. As can be seen from the gene pairs identified to date, one gene can be in more than one pair.

Results

**[0080]** Reciprocal Patterns of BDNF and BDNFOS Expression in Electrically Active Human Brain.

**[0081]** FIG. **1***a* shows a table of the 7 patients used for the described studies together with quantified in vivo spike frequencies, tissue locations, and pathological descriptions. Patients varied both in sex and age, but were chosen because of the availability of both high and low interictal spiking neocortical brain samples from nearby brain regions for each patient that were removed as part of their seizure surgery treatment. FIG. **1***b* shows how each of these pairs was selected with a short sample of the in vivo EEG recording that illustrates relative difference in interictal spiking. It is important to emphasize that because of genetic differences, medication effects, and effects of tissue processing the described internally controlled experimental design is crucial. Although patients are listed with different pathological diagnoses from multiple neocortical regions, only tissue samples that showed a normal cortical architecture were used so as not to influence the major variable of interest: increased brain activity.

**[0082]** Because of the potential regulatory relationship of transcripts that code for BDNF with those that encode the partially antisense BDNFOS the relative expression levels of BDNF and BDNFOS between paired high/low spiking regions of human neocortex using quantitative real-time PCR (qPCR) for each patient (FIG. **1***c*) was compared as a first step. In most patients, BDNF expression was higher in more electrically active regions, whereas BDNFOS lncRNA levels were significantly reduced in the high spiking regions. EGR1 expression was used as a positive control for high spiking human cortical brain regions as its expression has been shown to be directly proportionate to interictal brain activity. These results suggest that increased BNDF levels could in part be regulated by a decrease in antisense-binding BDNFOS.

**[0083]** BDNFOS is a Negative Regulator of BDNF in an In Vitro Human Cell Culture System.

**[0084]** The genomic antisense orientation of BDNF and BDNFOS is shown in FIG. **2***a*, where both overlapping and non-overlapping regions are delineated. Perturbation of lncRNA levels at multiple cis-antisense lncRNA/mRNA pairs affects levels of the cognate mRNAs. To distinguish whether the lncRNA BDNFOS directly regulates BDNF mRNA levels three siRNAs targeting human BDNFOS (FIG. **2***a*) were designed and used in qPCR to interrogate BDNFOS lncRNA and BDNF mRNA levels after the siRNA transfections. BDNFOS siRNAs were individually transfected into the human neuroblastoma cell line SH-SY5Y by electroporation, and caused reproducible BDNFOS knockdown at 24 h (all 3 siRNAs) and 48 h (only S2). Two of the siRNAs led to knockdown of BDNFOS by over 70% (FIG. **2***b*). BDNFOS knockdown by these dsRNAs consistently led to increased BDNF mRNA levels (between 1.5- and 3.5-fold change), suggesting that the cis-antisense BDNFOS RNA functions as a negative regulator of human BDNF (FIG. **2***b*).

**[0085]** Transcriptome-Wide Profiling of Human Protein-Coding and Long Non-Coding RNAs Reveals Activity-Dependent Gene Pairs.

**[0086]** The functional validation of the primate-specific BDNF/BDNOS pair suggests the potential for many more lncRNA/mRNA and/or lncRNA/lncRNA regulatory relationships across the human genome that may vary as a function of brain activity. The same paired RNA samples from the same 7 patients were thus used to identify members of the activity-dependent coding/noncoding and noncoding/noncoding interactome.

**[0087]** 4,004 mRNAs from the catalog array (1,944 upregulated and 2,060 downregulated in high-activity areas) were identified using the provided criterion for differential expression. On the lncRNA arrays, 86 of the 111 positive control genes were upregulated, and 1,288 lncRNA genes were differentially expressed between high-activity and low-activity neocortical regions (698 upregulated lncRNA genes and 590 downregulated lncRNA genes in high-activity areas). BDNF was represented on both the coding microarray and, as a brain-expressed known control gene, on the lncRNA microarray. BDNF was upregulated in high-activity tissue

from all 7 patients according to both array platforms: coding microarray, median 3.6-fold change; lncRNA microarray, median 2.8-fold change.

**[0088]** Of the lncRNAs identified to be differentially expressed at high-activity regions, 290 were found at genomic loci which the described genomewide analysis of UCSC cDNA-to-genome and EST-to genome alignments revealed to contain sense-antisense overlaps. At least 4 of these mRNA/lncRNA cis-antisense pairs were co-differentially expressed in all 7 patients (FIG. **3***b*). Only one of the 4 pairs (BDNFOS/BDNF) identified to date featured an inverse differential expression profile. The other 3 pairs all had a positive, direct-correlation. These pairs include lncRNAs cis-antisense to MAPK1 IP1L (MAP Kinase 1 Interacting Protein 1-like, potentially a modulator of MAP Kinase 1, whose role centers on the CREB activation pathway upstream of brain activity-dependent gene expression); PURB (purine-rich element binding protein, a gene expression regulator); and C11ORF96, a human homolog of the rat AG2 gene, induced as a consequence of sustained long-term potentiation in vivo in rat hippocampus and therefore implicated in neuronal plasticity. In summary, the mRNAs of at least 3 of the 4 co-differentially-expressed LncRNA/mRNA cis-antisense pairs have neuronal functions centered on synaptic plasticity.

**[0089]** A set of 276 identifed lncRNAs differentially expressed at high-activity brain regions reside at genomic loci corresponding to some of the 808 human mRNA/lncRNA neighbor-gene pairs in which a protein-coding gene and an lncRNA gene were non-overlapping but encoded within 10 kb of each other along the genome (Jia et al. 2010, incorporated by reference for its teachings regarding the same). However, initially only 4 mRNA/lncRNA neighbor gene pairs were identified as co-differentially expressed in the group-wise analysis of the 7 patients (FIG. **3***b*). These 4 co-differentially-expressed neighbor-gene pairs contained lncRNA genes neighboring the mRNAs ARC (activity-regulated cytoskeleton-associated), a key regulator of neuronal receptor endocytosis required for both synaptic plasticity and long-term memory, L-plastin, relevant to the activity-dependent MAPK/CREB activation by its placement within the human MAPK interactome, SMEK2, a regulatory subunit of Ser/Thr phosphatase 4, and CYR61, a secreted protein that associates with the extracellular matrix and the cell surface, regulates Akt activation, and is differentially expressed in autism.

**[0090]** 1,288 lncRNA genes were determined by our custom microarray to be differentially expressed at high-activity areas of the human neocortex. Some of these, including the lncRNA MALAT-1 which is a regulator of several synaptic genes, are indispensable components of specific nuclear bodies, while other lncRNAs regulate imprinting genes and still others perform essential catalytic roles. Differential expression of at least five of these known nuclear RNAs (FIG. **3***b*, bottom) was significant. MIAT, the sole member of this group which was downregulated in the more active areas, delineates a neuronal nuclear domain and was shown to be both a direct target and a putative co-activator of the transcription factor Oct4. In contrast, the levels of the other four known nuclear lncRNAs were increased in the more active areas. These lncRNAs included: KCNQ1OT1, which may regulate imprinting by recruiting the DNA methyltransferase DNMT1 to differentially methylated regions; RPPH1, the catalytic-RNA component of RNase P, essential for tRNA 5'-end maturation and for regulating Pol III-dependent tRNA transcription; NEAT1, an essential component of nuclear paraspeckles which suppresses the nucleocytoplasmic export of Alu-containing RNAs; and NEAT2 (MALAT-1), an essential component of nuclear speckles and a regulator of synaptic genes.

**[0091]** A second unbiased approach was also used to identify activity-dependent lncRNAs with potential importance in synaptic plasticity transcriptional regulatory networks. A number of coding genes including EGR1, EGR2, FOS, and DUSP6 are expressed in human brain in direct relation to the degree of epileptic discharges. Using co-expression clustering of mRNAs, these and other genes that have the same expression pattern across the seven patients were identified. Further, lncRNAs whose pattern of expression correlated with this group of coding genes were identified. FIG. **4**, constructed from our coding/non-coding transcriptome quantitation integration by Cytoscape software (Smoot et al. 2011, incorporated by reference for its teachings regarding the same) illustrates co-expression of 26 differentially expressed lncRNA genes with 13 differentially expressed mRNAs. In this diagram, genes that are closer together are more tightly linked. While it appears that there are two linked clusters of coding versus non-coding genes, this apparent separation may in fact be due to the overall lower level of expression of the lncRNAs compared to the mRNAs (by a factor of almost 80-fold). Some of these lncRNAs, such as NEAT1, are not at or near known coding-gene loci (FIG. **3**), while others are adjacent to known genes that are not differentially expressed. IL8RBP, a complex mosaic lncRNA transcript that combines unique upstream exons with an IL8RB pseudogene downstream exon, is differentially expressed, although its parental gene IL8RB, and the RUFY4 known gene which the pseudogene overlaps, are not detectable above background in the same samples on our protein-coding gene arrays.

**[0092]** Time-Dependent Patterns of lncRNA/mRNA Co-Differential Expression with Chronic Depolarization of Cultured Human Neuronal Cells.

**[0093]** To facilitate the study of primate-specific activity-dependent, lncRNA/mRNA and/or lncRNA/lncRNA gene pairs, an in vitro system of repeated depolarization using the human SH-sy5y neuroblastoma cell line was developed. FIG. **5***a* shows that while a single treatment of these cells with 100 mM KCl leads to transient CREB activation (phosphor-CREB), repeated 5 min exposures with KCl separated by 2-h intervals lead to more sustained CREB activation, similar to that observed in highly spiking human neocortex and in an animal model of frequent interictal spiking (FIG. **5***b*).

**[0094]** Gene expression over a 24 h time course together with BDNF and BDNFOS expression by qPCR was compared. Repeated depolarization led to a marked and more sustained increase in EGR1 activation (a marker of epileptic activity in the brain) (FIG. **5***b*, bottom panels). Even though EGR1 goes up within 4 h, both BDNF and BDNFOS were initially downregulated. However, whereas BDNF and BDNFOS almost return to baseline levels 24 h after a single depolarization, cultures that were repeatedly depolarized showed a small but significant increase in BDNF expression, while BDNFOS remained downregulated. The chronic depolarization-induced reciprocal BDNFOS depletion and BDNF mRNA increase in culture parallels the inverse relationship between BDNFOS and BDNF in high-activity areas of the human brain shown in FIG. **1**. Therefore, this cell culture system was applied to interrogate the activity-dependent expression patterns of other lncRNAs and their cis-encoded partner mRNAs from FIG. **3**.

**[0095]** In FIG. **5***c*, chronic depolarization of SH-sy5y cells revealed a number of distinct time-dependent patterns. Unlike the cis-antisense BDNF/BDNFOS pair, the AF086035/MAPK1P1L cis-antisense pair had an opposite effect of decreasing lncRNA and increasing mRNA levels early in the timecourse, although by 24 h the lncRNA level showed a slight increase. The BC013641/ARC neighbor-gene pair displayed increased expression of both genes at the 4 h timepoint in the depolarization-treated Sh-sy5y cells, mirroring the increased expression of both genes in the high-activity human brain. At subsequent time points, ARC mRNA levels decreased back to the pre-treatment levels although a sustained elevated level of the BC013641 lncRNA encoded near the ARC gene along the genome (FIG. **5***c*) was observed. Since the BC013641 gene is located approximately 6 kb from ARC with a divergent genomic orientation relative to ARC, two custom siRNA oligonucleotides targeting BC013641 were designed. Although both siRNAs knocked down BC013641, only one led to a 1.25-fold increase in the mRNA level of the neighboring ARC gene, suggesting that reciprocal gene expression directionality at lncRNA/mRNA pairs may occur at neighbor-gene loci such as BC013641/ARC, and not solely at sense-antisense loci such as BDNFOS/BDNF (data not shown). The BC047792/PURB cis-antisense pair showed increased expression of both genes, which mirrored the coordinate increase observed in high-spiking brain regions, however, in contrast to BC013641/ARC, expression of both transcripts were maximal at the 8h time point and returned to baseline at 24 hours, showing no sustained increase in lncRNA expression. Finally, the time course of one lncRNA, NEAT1, that has a potentially far-reaching regulatory role was examined. NEAT1 goes up within 4 h, returns to baseline at 8 h, but shows some chronic elevated expression at 24 h. NEAT1 is not in a cis-encoded pair, but it is in a trans-encoded network as one of the two targets of RPPH1, another lncRNA. Therefore, RPPH1/NEAT1 are a trans-encoded lncRNA/lncRNA pair. RPPH1/NEAT2 (synonym of NEAT2 is MALAT-1) are another trans-encoded lncRNA/lncRNA pair.

**[0096]** Unless otherwise indicated, all numbers expressing quantities of ingredients, properties such as molecular weight, reaction conditions, and so forth used in the specification and claims are to be understood as being modified in all instances by the term "about." Accordingly, unless indicated to the contrary, the numerical parameters set forth in the specification and attached claims are approximations that may vary depending upon the desired properties sought to be obtained by the present invention. At the very least, and not as an attempt to limit the application of the doctrine of equivalents to the scope of the claims, each numerical parameter should at least be construed in light of the number of reported significant digits and by applying ordinary rounding techniques. Notwithstanding that the numerical ranges and parameters setting forth the broad scope of the invention are approximations, the numerical values set forth in the specific examples are reported as precisely as possible. Any numerical value, however, inherently contains certain errors necessarily resulting from the standard deviation found in their respective testing measurements.

**[0097]** The terms "a," "an," "the" and similar referents used in the context of describing the invention (especially in the context of the following claims) are to be construed to cover both the singular and the plural, unless otherwise indicated herein or clearly contradicted by context. Recitation of ranges of values herein is merely intended to serve as a shorthand method of referring individually to each separate value falling within the range. Unless otherwise indicated herein, each individual value is incorporated into the specification as if it were individually recited herein. All methods described herein can be performed in any suitable order unless otherwise indicated herein or otherwise clearly contradicted by context. The use of any and all examples, or exemplary language (e.g., "such as") provided herein is intended merely to better illuminate the invention and does not pose a limitation on the scope of the invention otherwise claimed. No language in the specification should be construed as indicating any non-claimed element essential to the practice of the invention.

**[0098]** Groupings of alternative elements or embodiments of the invention disclosed herein are not to be construed as limitations. Each group member may be referred to and claimed individually or in any combination with other members of the group or other elements found herein. It is anticipated that one or more members of a group may be included in, or deleted from, a group for reasons of convenience and/or patentability. When any such inclusion or deletion occurs, the specification is deemed to contain the group as modified thus fulfilling the written description of all Markush groups used in the appended claims.

**[0099]** Certain embodiments of this invention are described herein, including the best mode known to the inventors for carrying out the invention. Of course, variations on these described embodiments will become apparent to those of ordinary skill in the art upon reading the foregoing description. The inventor expects skilled artisans to employ such variations as appropriate, and the inventors intend for the invention to be practiced otherwise than specifically described herein. Accordingly, this invention includes all modifications and equivalents of the subject matter recited in the claims appended hereto as permitted by applicable law. Moreover, any combination of the above described elements in all possible variations thereof is encompassed by the invention unless otherwise indicated herein or otherwise clearly contradicted by context.

**[0100]** Specific embodiments disclosed herein may be further limited in the claims using consisting of or consisting essentially of language. When used in the claims, whether as filed or added per amendment, the transition term "consisting of" excludes any element, step, or ingredient not specified in the claims. The transition term "consisting essentially of" limits the scope of a claim to the specified materials or steps and those that do not materially affect the basic and novel characteristic(s). Embodiments of the invention so claimed are inherently or expressly described and enabled herein.

**[0101]** Furthermore, numerous references have been made to patents and printed publications throughout this specification. Each of the above-cited references and printed publications are individually incorporated herein by reference in their entirety.

**[0102]** In closing, it is to be understood that the embodiments of the invention disclosed herein are illustrative of the principles of the present invention. Other modifications that may be employed are within the scope of the invention. Thus, by way of example, but not of limitation, alternative configurations of the present invention may be utilized in accordance with the teachings herein. Accordingly, the present invention is not limited to that precisely as shown and described.

REFERENCES

**[0103]** BARKMEIER D T, AGARWAL R, ATKINSON M, FLANAGAN D, SHAH A K, JAFARI-KHOUZANI K, AND LOEB J A, (2012) High inter-reviewer variability of spike detection on intracranial EEG addressed by an automated multi-channel algorithm, Clin Neurophysiol, 123 (6):1088-95.

**[0104]** BATES, D., M. MAECHLER AND B. BOLKER, 2009 lme4: Linear mexed-effects models using S4 classes, pp. R Foundation for Statistical Computing, https://r-forge.r-project.org/R/?group_id=60.

**[0105]** BEAUMONT, T., B. YAO, A. SHAH AND J. LOEB, in revision 2011 Layer-specific CREB target gene induction in human neocortical epilepsy.

**[0106]** GROSS, J., 2006 nortest: Tests for Normality. R package version 1.0., pp. R Foundation for Statistical Computing, http://cran.r-.project.org/web/packages/nortest/.

**[0107]** HOLLOWAY, M. G., G. D. MILES, A. A. DOMBKOWSKI and D. J. WAXMAN, 2008 Liver-specific hepatocyte nuclear factor-4alpha deficiency: greater impact on gene expression in male than in female mouse liver. Mol Endocrinol 22: 1274-1286.

**[0108]** JIA, H., M. OSAK, G. K. BOGU, L. W. STANTON, R. JOHNSON et al., 2010 Genome-wide computational identification and manual annotation of human long noncoding RNA genes. RNA 16: 1478-1487.

**[0109]** JIN, W., R. M. RILEY, R. D. WOLFINGER, K. P. WHITE, G. PASSADOR-GURGEL et al., 2001 The contributions of sex, genotype and age to transcriptional variance in *Drosophila melanogaster*. Nat Genet 29: 389-395.

**[0110]** KERR, M. K., and G. A. CHURCHILL, 2001a Experimental design for gene expression microarrays. Biostatistics 2: 183-201.

**[0111]** KERR, M. K., and G. A. CHURCHILL, 2001b Statistical design and the analysis of gene expression microarray data. Genet Res 77: 123-128.

**[0112]** KERR, M. K., M. MARTIN and G. A. CHURCHILL, 2000 Analysis of variance for gene expression microarray data. J Comput Biol 7: 819-837.

**[0113]** LITTELL, R., G. MILLIKEN, W. STROUP, AND R. WOLFINGER, 1996 *SAS System for Mixed Models*. SAS Institute, North Carolina, USA.

**[0114]** LOEB, J. A., 2010 A human systems biology approach to discover new drug targets in epilepsy. Epilepsia 51 Suppl 3: 171-177.

**[0115]** LOEB, J. A., AND T. L. BEAUMONT, 2009 What goes in is what comes out: The design and analysis of microarray experiments, pp. 213 in *Bioinformatics for Systems Biology*, edited by S. KRAWETZ. Humana Press, New York, N.Y.

**[0116]** R DEVELOPMENT CORE TEAM, 2010 R: A language and environment for statistical computing., pp. R Foundation for Statistical Computing, http://www.r-project.org/, Vienna, Austria.

**[0117]** RAKHADE, S. N., B. YAO, S. AHMED, E. ASANO, T. L. BEAUMONT et al., 2005 A common pattern of persistent gene activation in human neocortical epileptic foci. Ann Neurol 58: 736-747.

**[0118]** SMOOT, M. E., K. ONO, J. RUSCHEINSKI, P. L. WANG and T. IDEKER, 2011 Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics 27: 431-432.

**[0119]** SMYTH, G. K., and T. SPEED, 2003 Normalization of cDNA microarray data. Methods 31: 265-273.

**[0120]** SONE, M., T. HAYASHI, H. TARUI, K. AGATA, M. TAKEICHI et al., 2007 The mRNA-like noncoding RNA Gomafu constitutes a novel nuclear domain in a subset of neurons. J Cell Sci 120: 2498-2506.

**[0121]** STRIMMER, K., 2009 fdrtool: Estimation and Control of (Local) False Discovery Rates. R package version 1.2.6, pp. R Foundation for Statistical Computing, http://strimmerlab.org/software/fdrtool/.

**[0122]** TANAKA, T. S., S. A. JARADAT, M. K. LIM, G. J. KARGUL, X. WANG et al., 2000 Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray. Proc Natl Acad Sci USA 97: 9127-9132.

**[0123]** WOLFINGER, R. D., G. GIBSON, E. D. WOLFINGER, L. BENNETT, H. HAMADEH et al., 2001 Assessing gene significance from cDNA microarray expression data via mixed models. J Comput Biol 8: 625-637.

**[0124]** YAO, B., S. N. RAKHADE, Q. LI, S. AHMED, R. KRAUSS et al., 2004 Accuracy of cDNA microarray methods to detect small gene expression changes induced by neuregulin on breast epithelial cells. BMC Bioinformatics 5: 99.

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 36

<210> SEQ ID NO 1
<211> LENGTH: 2059
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 1

atcgcgagat caggaaggtg gccgagtgtg tcgccgcggc catcaggcac ttctccttcc 60

tgcccttgta tgaagaagga tgtgtttgct tccccttgtg ccatgattgt aaatttcctg 120

aggcctcctc agccctgcgg aactggctag agcaatgtat cttaggctca cttaaggaag 180

ctgtagagat gagcccaagg agggaaacca gaagagcccc ccaggctcac cagttgtttg 240

ttggctccct acaaacatgt cattcaagtg gctaatctta caacagcaca aattcatcta 300

-continued

accagaaaga gaagaggagg ctccaaaggc acttgactac tgagcatcac cctggacgtg 360

tacaagtctg cgtccttatt gttttcttca ttgggccgaa ctttctggtc ctcatccaac 420

agctcttcta tcatgtgttc gaaagtgtca gccaatgatg tcaagcctct tgaacctgcc 480

ttgggcccat tcacgctctc cagagtccca tgggtccgca cacctggaga tactctatta 540

tagcaaagaa gaaagataat ttcattgagc catcctgttt tacagcaccc aacagaatcc 600

cttcaaagcc tcgtggtctg acaccctatg ctacgtgact tgtgacccat ccatttgtca 660

tgttcttcgg gaatgtggct aaggggctaa gatgtgactt gaaaagaaag gtagaacaag 720

atcatctcaa atttattatc aaggaatagt tcagaaaacg acttcagacc acagagacag 780

cagaacagat ggtccggcat ggatagagca tcagacactc acagactgtg ccaacaagag 840

ccatcgagtc aaaacagcca aaggaaggag ggtcatggaa tgggttctct cacaccaaac 900

tgatgcccag aggccctcag catgaataac aaaggcaacc agacccacaa gccatactga 960

gtggatacaa aacctatacc taagctgaca tcccaaatgt gtgtggcaag ttagatgatg 1020

atggcacaaa agacagaaca ccttgctttc tggccattgt cagctcttgg aagagagcac 1080

acttttagag gagcagctgc aaggaccctg agaacaaaac tggaaatgtc tgttatgaaa 1140

gccttcacag gaaattctgc aagtggcaac gtgggtccat tccgtgtgtg tcactagagc 1200

tggcgcaagc ccatggccat ggtgaggcag cgtttccact ggaactaatc tgatacctgc 1260

accagctctt gcaactgtgc agtgttccca ctgcaaacta cggatgggag aggataaaga 1320

acttcaatct ttaaaaaaga gaggattttc cctcctggtg agtcaaaatg aacaagaaat 1380

accccaggac ctcccttccc tccttggcca ttaatgagat gaaggcaatt aactcacata 1440

gtataaatga atcatttgag gtgatgactg cattttaggc aaatgatgac tttcttggtt 1500

ccattggttt gcaagtaaaa gttacacaca ttgaaaagac actgaaacag atttcctaaa 1560

tgcttcattt tctggatgca ccaatgttga cctactatac atgttaaatg gttttaaaat 1620

atcaccttaa aataaaggaa acttccagct actaactcag ctctgaatgg gctatgaaag 1680

gctccaaagg tatgtgaaaa attactgtta ttttgcttta aaaaatgtga tgtctaagag 1740

tgtctgcaat gttctaatgc ttcaaaacat gtacgtaagc cttgtttatc tggaaatcat 1800

ttctttctgc ttatatcatt tataaataga aaatgttctg taataactta aaatagttcc 1860

acatacataa tgcttttcgt gtcataatac ttactactgg tctatattta ccaacattta 1920

tcacatttta caaaatgaag tagaagaaaa aaaagacaac gactttatgg ccctggaatt 1980

ccagtaatgg tgaccaacat gttttaaatt ccagtaaagg ttatggttac atttcaaaaa 2040

aaaaaaaaaa aaaaaaaaa 2059

<210> SEQ ID NO 2
<211> LENGTH: 4755
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 2

cacacacaca cacacacaca gagagaacat ctctagtaaa aagaaaagtt gagctttctt 60

agctagatgt gtgtattagc cagaaaaagc caaggagtga agggttttag agaactggag 120

gagataaagt ggagtctgca tatgggaggc atttgaaatg gacttaaatg tcttttaat 180

gctgactttt tcagttttct ccttaccaga cacattgttt tcatgacatt agccccaggc 240

-continued

atagacacat cattaaaatg aacatgtcaa aaaatgattt ctgtttagaa ataagcaaaa 300

cattttcagt tgtgaccacc caggtgtaga ataaagaaca gtggaattgg gagccctgag 360

ttctaacata aactttcttc atgacataag gcaagtcttc tatggccttt ggtttcctta 420

cctgtaaaac aggatggctc aatgaaatta tctttcttct ttgctataat agagtatctc 480

tgtgggaaga ggaaaaaaaa agtcaattta aaggctcctt atagttcccc aactgctgtt 540

ttattgtgct attcatgcct agacatcaca tagctagaaa ggcccatcag acccctcagg 600

ccactgctgt tcctgtcaca cattcctgca aaggaccatg ttgctaactt gaaaaaaatt 660

actattaatt acacttgcag ttgttgctta gtaacattta tgattttgtg tttctcgtga 720

cagcatgagc agagatcatt aaaaattaaa cttacaaagc tgctaaagtg ggaagaagga 780

gaacttgaag ccacaatttt tgcacttgct tagaagccat ctaatctcag gttatatgct 840

agatcttggg ggcaaacact gcatgtctct ggtttatatt aaaccacata cagcacacta 900

ctgacactga tttgtgtctg gtgcagctgg agtttatcac caagacataa aaaaaccttg 960

accctgcaga atggcctgga attacaatca gatgggccac atggcatccc ggtgaaagaa 1020

agccctaacc agttttctgt cttgtttctg ctttctccct acagttccac caggtgagaa 1080

gagtgatgac catccttttc cttactatgg ttatttcata ctttggttgc atgaaggctg 1140

cccccatgaa agaagcaaac atccgaggac aaggtggctt ggcctaccca ggtgtgcgga 1200

cccatgggac tctggagagc gtgaatgggc ccaaggcagg ttcaagaggc ttgacatcat 1260

tggctgacac tttcgaacac gtgatagaag agctgttgga tgaggaccag aaagttcggc 1320

ccaatgaaga aaacaataag gacgcagact tgtacacgtc cagggtgatg ctcagtagtc 1380

aagtgccttt ggagcctcct cttctctttc tgctggagga atacaaaaat tacctagatg 1440

ctgcaaacat gtccatgagg gtccggcgcc actctgaccc tgcccgccga ggggagctga 1500

gcgtgtgtga cagtattagt gagtgggtaa cggcggcaga caaaaagact gcagtggaca 1560

tgtcgggcgg gacggtcaca gtccttgaaa aggtccctgt atcaaaaggc caactgaagc 1620

aatacttcta cgagaccaag tgcaatccca tgggttacac aaaagaaggc tgcaggggca 1680

tagacaaaag gcattggaac tcccagtgcc gaactaccca gtcgtacgtg cgggccctta 1740

ccatggatag caaaaagaga attggctggc gattcataag gatagacact tcttgtgtat 1800

gtacattgac cattaaaagg ggaagatagt ggatttatgt tgtatagatt agattatatt 1860

gagacaaaaa ttatctattt gtatatatac ataacagggt aaattattca gttaagaaaa 1920

aaataatttt atgaactgca tgtataaatg aagtttatac agtacagtgg ttctacaatc 1980

tatttattgg acatgtccat gaccagaagg gaaacagtca tttgcgcaca acttaaaaag 2040

tctgcattac attccttgat aatgttgtgg tttgttgccg ttgccaagaa ctgaaaacat 2100

aaaaagttaa aaaaaataat aaattgcatg ctgctttaat tgtgaattga taataaactg 2160

tcctctttca gaaaacagaa aaaaacacac acacacacaa caaaaatttg aaccaaaaca 2220

ttccgtttac attttagaca gtaagtatct tcgttcttgt tagtactata tctgttttac 2280

tgcttttaac ttctgatagc gttggaatta aaacaatgtc aaggtgctgt tgtcattgct 2340

ttactggctt aggggatggg ggatgggggg tatatttttg tttgttttgt gttttttttt 2400

cgtttgtttg ttttgttttt tagttcccac agggagtaga gatggggaaa gaattcctac 2460

aatatatatt ctggctgata aaagatacat ttgtatgttg tgaagatgtt tgcaatatcg 2520

-continued

atcagatgac tagaaagtga ataaaaatta aggcaactga acaaaaaaat gctcacactc 2580

cacatcccgt gatgcacctc ccaggccccg ctcattcttt gggcgttggt cagagtaagc 2640

tgcttttgac ggaaggacct atgtttgctc agaacacatt ctttcccccc ctccccctct 2700

ggtctcctct ttgttttgtt ttaaggaaga aaaatcagtt gcgcgttctg aaatatttta 2760

ccactgctgt gaacaagtga acacattgtg tcacatcatg acactcgtat aagcatggag 2820

aacagtgatt ttttttaga acagaaaaca acaaaaaata accccaaaat gaagattatt 2880

ttttatgagg agtgaacatt tgggtaaatc atggctaagc ttaaaaaaaa ctcatggtga 2940

ggcttaacaa tgtcttgtaa gcaaaaggta gagccctgta tcaacccaga aacacctaga 3000

tcagaacagg aatccacatt gccagtgaca tgagactgaa cagccaaatg gaggctatgt 3060

ggagttggca ttgcatttac cggcagtgcg ggaggaattt ctgagtggcc atcccaaggt 3120

ctaggtggag gtggggcatg gtatttgaga cattccaaaa cgaaggcctc tgaaggaccc 3180

ttcagaggtg gctctggaat gacatgtgtc aagctgcttg gacctcgtgc tttaagtgcc 3240

tacattatct aactgtgctc aagaggttct cgactggagg accacactca agccgactta 3300

tgcccaccat cccacctctg gataattttg cataaaattg gattagcctg gagcaggttg 3360

ggagccaaat gtggcatttg tgatcatgag attgatgcaa tgagatagaa gatgtttgct 3420

acctgaacac ttattgcttt gaaactagac ttgaggaaac cagggtttat cttttgagaa 3480

cttttggtaa gggaaaaggg aacaggaaaa gaaaccccaa actcaggccg aatgatcaag 3540

gggacccata ggaaatcttg tccagagaca agacttcggg aaggtgtctg gacattcaga 3600

acaccaagac ttgaaggtgc cttgctcaat ggaagaggcc aggacagagc tgacaaaatt 3660

ttgctcccca gtgaaggcca cagcaacctt ctgcccatcc tgtctgttca tggagagggt 3720

ccctgcctca cctctgccat tttgggttag gagaagtcaa gttgggagcc tgaaatagtg 3780

gttcttggaa aaatggatcc ccagtgaaaa ctagagctct aagcccattc agcccatttc 3840

acacctgaaa atgttagtga tcaccacttg gaccagcatc cttaagtatc agaaagcccc 3900

aagcaattgc tgcatcttag tagggtgagg gataagcaaa agaggatgtt caccataacc 3960

caggaatgaa gataccatca gcaaagaatt tcaatttgtt cagtctttca tttagagcta 4020

gtctttcaca gtaccatctg aatacctctt tgaaagaagg aagactttac gtagtgtaga 4080

tttgttttgt gttgtttgaa aatattatct ttgtaattat ttttaatatg taaggaatgc 4140

ttggaatatc tgctatatgt caactttatg cagcttcctt ttgagggaca aatttaaaac 4200

aaacaacccc ccatcacaaa cttaaaggat tgcaagggcc agatctgtta agtggtttca 4260

taggagacac atccagcaat tgtgtggtca gtggctcttt tacccaataa gatacatcac 4320

agtcacatgc ttgatggttt atgttgacct aagatttatt ttgttaaaat ctctctctgt 4380

tgtgttcgtt cttgttctgt tttgttttgt tttttaaagt cttgctgtgg tctctttgtg 4440

gcagaagtgt ttcatgcatg gcagcaggcc tgttgctttt ttatggcgat tcccattgaa 4500

aatgtaagta aatgtctgtg gccttgttct ctctatggta aagatattat tcaccatgta 4560

aaacaaaaaa caatatttat tgtattttag tatatttata taattatgtt attgaaaaaa 4620

attggcatta aaacttaacc gcatcagaac ctattgtaaa tacaagttct atttaagtgt 4680

actaattaac atataatata tgttttaaat atagaatttt taatgttttt aaatatattt 4740

tcaaagtaca taaaa 4755

-continued

<210> SEQ ID NO 3
<211> LENGTH: 361
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 3

tttttttttt tcagattttc ctatgtattt ttatttttttc cagatcgtta aaaaaaaaaa 60

ttaatgtgtt agcattctgg ttggaaatgt attatattta cagattaatc taggggaaca 120

tctatttaaa atgtcaaatc atctcgttca agaacaaggt atagtatagt ccctcggtac 180

ctgagggaga ttggttctag gacccccata gatactcaaa tcccaggatg ctcaagtacc 240

tgatataaaa tggtgtagta tttgcatata acctatgcac atcttcccgt ataaatcatc 300

tctatattat ttataatact taaaagtgtg aatactatat aattatatta acttatattt 360

t 361

<210> SEQ ID NO 4
<211> LENGTH: 6469
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 4

ggcgcttcct gttccggcgc caggaggagc cgcgcgctgc tggtgctgtt gccgccgctg 60

ctctagctgc cgtcagtcag gctgcgcccg cgtcttcagg gcccagtccc tcggacccat 120

cgccgcttct agaccctact gcggtctcgg atattgccgg gaaaatgtct gatgaatttt 180

cgttggcaga tgcactacct gaacactccc ctgccaaaac ctctgctgtg agcaatacaa 240

aacctggcca acctcctcaa ggctggccag gctccaaccc ttggaataat ccgagtgctc 300

catcttcagt gccatctgga ctcccaccaa gtgcaacacc ctccactgtg ccttttggac 360

cagcaccaac aggaatgtat ccctccgtgc ctcccaccgg accacctcca ggacccccag 420

caccctttcc tccttccgga ccatcatgtc ccccacctgg tggtccttat ccagccccaa 480

ctgtgccggg ccctggcccc acagggccat atcctacacc aaatatgccc tttccagagc 540

tacccagacc atatggtgca cccacagatc cagctgcagc tggtccttta ggtccatggg 600

gatccatgtc ttctggacct tgggcgccag gaatgggagg gcagtatcct acccctaata 660

tgccatatcc atctccaggc ccatatcccg ctcctcctcc tccccaagcc cctggggcag 720

caccacctgt tccatggggc accgttccac caggagcctg gggaccacca gcaccatatc 780

ctgcccctac aggatcgtat cccacaccag gactctatcc tactcccagt aatcctttcc 840

aagtgccttc aggaccttct ggtgctccac caatgcctgg tggcccccat tcttaccatt 900

aagttaacaa tggacgaaga gatgacgctt tgctttttga agtacatgta tatgcacatg 960

aatgcatata taaaaattgc tggtttcact attagagggc attcatgaaa gaacaactct 1020

tgcacctctc agagaagata actgcctctt gtacttggat gcgtagtaca tcatatgtat 1080

acaatcagat aaaagcatag aagtaaatca ttcggatgtg atttttattt ggttttcatg 1140

gaaagttaaa gtgataaagt atattgaata gttctttgac agaatttgtt taaactatga 1200

aactacacac ttaaaaatct aagatgtgga ttattgttag aatctgcaac ttcattggca 1260

aattatttca agtatttttc tataatcact ttccccttct aaataaataa acttcgagaa 1320

taacccatca taatccaaac aaatgatgcc tcaacatttt gagctgctct gtcggacaaa 1380

-continued

taaacctggt cctcttgagg ttatattttg gatatacatt tttaaactgt cagtaattat 1440

tgtcagatgt ggagttcaat agccagccag tgttcatttt tatccttgag cttttagtaa 1500

aaacttcctg gttttatttt tagtcattgg gtcatacagc actaaagtct gctatttatg 1560

gaaactaact tttttgtttt taatccaggc caacatgtat gtaaattaaa tttttagata 1620

attgattatc tctttgtact acttgagatt tgattatgag atgtgcatat tgctttggga 1680

agagctcgag gaaggaaata attctctcct ttgttttgaa cctcaaacta gataaaccct 1740

aggaattgct taactgcaac aagtaatttt cattcccaca aaaacctgag gcagctcttt 1800

tgcccagagc gttccctgta gccacccccа ccccacttgc ccttggttct ttagaaggag 1860

cacacacatc ccttgattcc tccctgatgt ggtaaactgg cacactccag gggtctaaaa 1920

cataaaacag ttgtgtttag ggaaccttaa gtcatgcaga catgactgtt ctctttgtac 1980

aagtgtgaat caaaatatgt atctcttttt cagagtctgg ttaagctatg tcattgtcta 2040

ctgcatagtt tcctgagtct gtttgtaaag tgcttatggc taacagttca gttctgtatt 2100

tgttgacagg taaataagtg gagttgagtg ccatctttga aaaaattacc ctctagctct 2160

aacactgaaa ataataataa attgtagatc tctgcaacta agtttaaagc agtgtgactg 2220

tgttgcttaa atatcaagta ttgtttataa ccaccaaaaa aaaaaagccc tggtagtttt 2280

ttggcacctt atgtttaaat cagattctta gatttggagt agacctgacc ttgttattta 2340

ttagataaca ttttgaatgt atccattgga tttctaaaat gtattgtgaa tttctcagac 2400

aaacaggatt tatgctggag ctctgttttg cttagaaata aaatatttag tagtttattt 2460

ctgctctaat taaaatgtca agaatgccaa atgctgccag ttttttggtt tgatagctac 2520

ctccttctaa gaaagcaaaa tggttacctt tgagaggaac attcagtgtt taatcatccc 2580

ttatgttaac tagatgatag attcaagctt ttagaaatga gaaagtagaa actaatttgt 2640

taagatattt tcagactgcg gaatgttgtt agctttttct ttcacttctc ttcaaggaca 2700

ggtgttagct gtctacaata ctgttgaact ctgttgtcaa agtagccccc ttagtctaca 2760

aggcaggtag ccttggcttg aattatcaat atcaaaatgt cagttaacca tggagggata 2820

aagtaatgtg aaaagtgaga tggctgcaaa gatagctctc cttacagtta ttttggctgt 2880

cctacattgg gataagctga caaattagca gtatttagtt taacactgga gcaaatataa 2940

tttgagtagg aagaagagat agcaggtttg ggaatctata attatgaagt ccattgattt 3000

tgggagaaaa tctgttgcta aaggatttga agggccatga acacaatttg ggattattac 3060

tccctataag tataataatt ttgctagtga cccatactgt ccagtgtgcc ctaaatcata 3120

ctgctattgt actccctttg ttttcaagga ctttgcaact ggtatttggg ggagattttt 3180

ttttttttt gagacggagt ctcgctctgt cgcccatgct ggagtgcagt ggtgctatct 3240

tggttcactg caagctccat ctcccaggtt cacaccattc tcctgcctca gcctcccaag 3300

cagctgggac tacaggtgcc cgccaccatg cccggctaat ttttttttt ttttttttt 3360

agtagagatg gggtttcact gtgttagcca ggatggtctc gatctcctga cctcgtgatc 3420

tgcccgcctt ggcttcccaa agtgctggga ttacaggcgt gagccaccac gtccggccga 3480

ttttttttt ttttttaat gtaagaatgg agataaaagg gataatataa tttgctttta 3540

tattgttatt tttgtaaagc atctttttctt caattcttgt tggcattctg ggccaaaata 3600

tttcaggttg gttcggtgtg gagttaagaa aagcaggcgt tttagtggag aaatggggaa 3660

-continued

cagcatcaag aaaggctttt ttcctttttt ctttttttt tggagacaga gtcttgccct 3720

gtcacccagg ctggagtgca atggcgtgat cttggctcgc tgcaacctct gcctccaggt 3780

tcaaacgatt cttctgcctc agcctcccaa gtagctggga ttacaggtgc ccgccaccac 3840

acccatttt gtatttttag tagagacggg ggtttcacca tgttggccag ggtggtctga 3900

aactcctgac ctcgtgatcc gcctgcctca gcctcccaaa gtgctgagat tacaggcgtg 3960

agccaccatg cgtgaccttt ttttcttttt aaaagggaac aatgttgctt tcaaaacaag 4020

acatgctagg ctgaaactga tttatggaaa agactgcttg ttagcaagta tatttggtct 4080

tgaggggat acagattata gaatatgctg acatttgggc ttcagaggaa gaattttcaa 4140

atctaatgga aatagttgag gtgttcagga atgctgtttc ttggagttgg aagcttaggt 4200

tttgaaatgt tgaaaccaaa aagacaaaaa ttaaaacata gaccttaggt cgtcattcac 4260

acccggttct caagaatcaa gtggagcact tcaaagacct tggcttgtct gtcccatcct 4320

gccactttct catcttttca tgcttttgaa gacaccattt acagctctga ctcagcccta 4380

ttttgtgtaa agtaatatat tgattattca gaaatagaca atacattttt taattaccca 4440

aggactgact gttttgtgca ttttactgtt ggttgtcttc agtagagaat agtaataggg 4500

cagagaaaag tatatatttt gcctcagtca gtcccaccac cacaatggac tattgggata 4560

ttttctaaaa aaccaatcaa tttgcccatg attacctcac aaataattag tgctacctgg 4620

ggtactctca aatatacagc ttttgaaact gtagatgaaa aaagctctac tcagagtttt 4680

tgtcaagact gtgcctgggt tgaatatcag tcaattgcct acacttctaa acaataagtg 4740

ccaatgtctc aattttctca ccctgaatga tagaagctag ctttatcaaa tgccaaggtt 4800

agaaagcctg gaaataaaac ttaagcacag acattcaagt ttttgaaaag cataagccta 4860

aattcagata aatcacactg atatattgta ctatgcatag aaagttgtag gtggcgttca 4920

gggaagactt tgattttaat aaagcaatat ttagtattga agacaaacac tttttatttt 4980

cagatttctg ccaagtaaaa cagaaattgc caataaaata atcagtattt tgtaaatggc 5040

aggcaagctt ctggctgtcg aaaacatctg agtcatttat tcagtagaca atatgtcctt 5100

gatccaggtt ctttgccagc tataagggaa tccctgtcct tgagaggctc atagtctata 5160

agtaacatta cagaatttgt tagcataccc attcattatt agttttacct aaacgtgtta 5220

ggatcactac tggtggaaat tgtaaccagc ctttgggcat cttaaagggt gacatgtggc 5280

atgccttttt ttttttttaa gaatttaatg tttttcaaga ttgtagtgtt gatcagcgca 5340

acaattcaag tgtgcaaagt aacaggatag tttgcctctt cactttaccc ctggataaag 5400

gcactttcac tgcctgtcac tgatcagcag atactgactt gttgccatta agtgaacttg 5460

acttcttatg tgtgctctat gagtttgttg taatttttctt cttgaaattg tgatttttca 5520

ctgacagtaa tgacaaattt aatgtatgta attgtctatg cattttaagt taaactgcct 5580

aaaatgtgat ttgagacata tacatatgtt tgtattataa attgtaagca atcagtttga 5640

gatactaggt tttatcacct gctgctgtat ttgtaaacaa agacaaatgt tgctttaaga 5700

agtaattata attaggaata ggctatggat gtgatacttg gtatttttta agataaactt 5760

gtttgctttt gtgtattata cctggaaact ttttttaaaa aatgtatttt catggtttca 5820

cagatttttc atgttatttt attctttagg cccaattctg ggcttctctg agcaagtcca 5880

gagcctaatt aactgtaaat ttgttgtcaa aaaggaagaa aaaagggcct gagatacctc 5940

-continued

tttgcatgtg acctgcattc actaaggata tctggaaacc acccttcctc cgcaaaccct 6000

ctcagcaaca tggtgtccat tgtggtgatt ttctcttctt ttaaggctag gctactcttg 6060

gtaaccagat tatccgtata tatgataata tgaagtcagg gaactttctc tgtctgtccc 6120

tactcccctc actcccccac tttctgttat gaaagatagt tctactttta tcattaactg 6180

ctacgcattt agtgagggtc acattattaa acttggagtt taccattttc ccacaggaga 6240

tttcgctggc attccttgga actcccaatt tcagtagggc aatgaatgaa tgaatacttt 6300

gcagtgctac ttttggaagg aatttctgct ttttgcctta tgattggaca aaatgcagct 6360

gtaaaatttt aaattgtttt tgatatgtta ttcaatatcc catgaaagta ttcacctaaa 6420

gtggagttat gaaatggatg gtgaaataat aagaccattc tggagcagg 6469

<210> SEQ ID NO 5
<211> LENGTH: 2042
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 5

agagctgctt gcttaaaaga ctctggcctc cctcttgctc cctctctggc catgtgacat 60

gcctgctccc cctccacctt ccaccatgag tggaagcttc tcaaggcctc accagaagaa 120

gatgttggtg ccatgcttgt atagcctgca gaactagata ggatttcacc atgttgccca 180

ggctggtctc gaactcctgg gctcaagcga tctgcccacc taggcctccc aaagtactgg 240

gattacaggc gttcactgcc cggctgagaa ggaggttctt ccgcgagaga aggctatgca 300

gaggggagga tctcaggcct gaagggtcta gaagcccact ggaaggtctt ccagacactt 360

gggacatcta gctcctgaag agtagaaaga ggagctgctg ccgaagaggt aagacagaga 420

gaggggcaca ggatgtggag tccgcccctc tgtcctcagg gagtttcttt ggagagagaa 480

gagaatgagg ctgggctcag aggaagttcc tcccactccg tggagctcaa actagtccag 540

tggtggcagc tgccctggct ggaataaagc atcgacccac agtggcctca catctctgag 600

gcctttgcag gttaagagcc tcctgccatg ataccccacc acagtggaac ccgttgttgg 660

gccagcgtgt tctcagggac tgggacacgg gccaaatact acagccacca aggccatagc 720

ctgtgctctc tctgagccga gggaagaaca ccctctccgg ccctgtagtt ccatttttgt 780

ttctttgttt catttcctct gctgtcttct ccagagctct gcgctgcagc tgagagactg 840

ggtatctcct agagattcct gacccccacg caccagcccc accaccagag ggcgcccgag 900

cacagagacg ccgcaaccgg agcaggagcc ctaaacctgt cagggtgggt cagccctgga 960

gtaatcctgg agggagggtc tgggagaggg ccccagtgtg gggacctgcc cggctcaacc 1020

aaatgtagtt tgttaccttg gtggccatct gcagccaggt ggccctgcca cactgctaaa 1080

ccagaggatt taggaagcag acagggattc gggaagtggc tgaagagcac tggaaagaca 1140

aactcctgcg gtgcaagagt cctggaccag gagtgcggag aactgggggc gctcctctcc 1200

ctgttaccct aggcttgttt tctcatatgt aaacaaaggc aggggcgggg agaagggaag 1260

gactcgatgg tacccaggtg cccttcaggc tctaaccaaa gacatgcccc accctcctgg 1320

aggaaccctg cggggcgcag agaacggctg gagacatagc ggccgcggtt tcccccgcag 1380

cgcaggagcg ccaggcactg cgctgagtca actgcggtca gggagtcacc tgcccgcgac 1440

agcgagcgcg cacccacact ctcgaccaac ccaggctggc ctccagccca ctctcgggtc 1500

-continued

ccaacgccca aacatctcgc attgctggct ccgacatctc cccacccgcc gccaggaatc 1560

agcgcagagg gagatgcgtc ttctgtggcg ggtgcggggc gaggaccctg accccagctc 1620

tgggaagtga actgtacgga ggagagagct ctccatcgtc agctctgcct ccgcccctcc 1680

tgcggcttca cttcagcccg gatgccggcg atggggaata gaccagggac gcaggagcac 1740

actggaccgg gggtggatcc acttggggag gcaggcaacc gacccgcgca ccggtggcgc 1800

aacctcgttc cggtgaccga agcttgtgga ggtagctcca ctcccggctc cctggtccga 1860

catccccctg ccacctcctc ccccacctcc ccgagtcttg ggagcctgga ccccgaaggt 1920

gcaaaaagtc aagagcctgg aaaactgcac tgtgcaaata tattgacttt atttgtctcc 1980

tttcaggagc ctcacagaca tatccaggta aaaagatcgt taaataaatg ccttcagcca 2040

tc 2042

<210> SEQ ID NO 6
<211> LENGTH: 1361
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 6

atggcattcg ctgtcatccg agctcagagc cgtgtgggca gccgcgggct atataagccg 60

cgagcctggc cgccgcgggg cagacggcga cagcagcggc ggcgagcgcc tcggagcgcg 120

gcggaacagc gccccccgag ccccgtgccc cccgacgggt ccgcccgccc gcccgccctc 180

ccgaggagcg ccggcccggg cccgcgaggg ccgccgccac cccgcagcag atttggatcc 240

cccgcccggc gagcccccgg ctgctgcctc ccgggggggcc ccggcgcagc ggccgccctc 300

ggagagcccc ggcgccccgc cgcccggccc cgcagacgcc ggaggcgcca tggccgccaa 360

gcccggcgag ctgatgggca tctgctccag ttaccaggcg gtgatgccgc acttcgtgtg 420

cctggccgac gagttcccgc agcccgtgcg gcccgccaag ctgcccaagg gccggggccg 480

gctgcggcgg ccgcgccagt ctcgcttcaa gacgcagccg gtgaccttcg acgagatcca 540

ggaggtggag gaggaggggg tgtcccccat ggaggaggag aaggccaaga agtcgttcct 600

gcagagcctg gagtgcctgc gccgcagcac gcagagcctg tcgctgcagc gggagcagct 660

cagcagctgc aaactgagga acagcctgga ctccagcgac tccgactcgg ccctgtaagg 720

ggcgccgccc gcgggggggga cgcgcgcgtc cgcggtccgc gcggggaccg gcgtgtgaac 780

cccgagagtg cccgcgccct gctcccgggg gacccgcaag gacccgggac cgccgctcct 840

cgcgcgctcg gactcccgcc ccgctgcgaa ccggtcggtg cgcccctcgc cgcgctcgcc 900

ctggcccggg agcgccggga gcggggccgc tttcctcgtc cttgtaaatg tttatttttt 960

aactcttccc agtgcgaact ctgctgtgag tgtgtgcggg gaggcgcgcc cgcgctgagt 1020

cggcggcggg tagccactcc atgcccttgt ccgatggttt gcaactccga ttttgcacac 1080

cgctccaccg tgccccccag cgcacaccca ttcacactca cgccaacact ctcgctgaac 1140

acttttataa ttgttaggcg tggccgttgg gactttgggc gcagcgcggc tgctactgcg 1200

tctggaggat tgatatttat ttttgcattg cgatggctga aggcatttat ttaacgatct 1260

ttttacctgg atatgtctgt gaggctcctg aaaggagaca aataaagtca atatatttgc 1320

acagtgcaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa a 1361

<210> SEQ ID NO 7

-continued

<211> LENGTH: 1475
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 7

gtagtacttg cggttctcac gcaccaagaa ttcgctcttg agcgcgcgcc gcggcccgcc 60

gccctcctcg gcgcccgccg ccagctgctc ggggctgcta gggcccagct gcgcgtagtg 120

ttctatgaag tcgcccagcg agtcgcggaa ctcggcggcc accgccatgg acagcgtgag 180

gcggctcttg gaaccgcccg cgcccacctc ggcgatcttg aggaagcggc ccttggcgtt 240

ctgcttcaca tctaagtaga agcgcttgtt ctggatgtcc agccgcttcg aggccagctc 300

ctgcgtctct tgctcgccgc cgccgcggga cgcgggctgg aacccgcacg gcccaccgcc 360

gccgccgcgc tcgctgccgc tgtcgccgtc cgccatcttc taggccgccg cctcgccgcc 420

accgcccgcc gcgctcgcgc ccccgccctc cggctcgcgc tccggggccc cccagcctcg 480

cccgcccggc tcgctcacgc gctcccgccg ctcatcgccg gccgccgccg ccccccccatc 540

gcctccgcgc cccgcccccg cgacttccgt gcagccctag ccacttccgg cgcgcgattt 600

cccgtgaagt cactccgtcc cgccctccgc ggcctcgcgg ccggaagcgc cgcggcggcg 660

tggcgcagag gggtgacgtc gggggccgcg ccgggcgggc ggcaagctga gagtaggcta 720

gtgttactgc tcggtttcta gccctgggga cttgccagcc ttagcgtgac tcagccgccc 780

gctgcgtctg tctacgcgcc tcttcccgtg acacttttac tttagactcc ggtgtcgtcc 840

acttccctcc ttctcgcttc cggggcgtta gctgagcata ctcaggcccc agcctgctcg 900

cgctggttgt gtgctgggaa cagagttcca gtctatatac tggcgaccgg ggggagggag 960

atcctttgga gcactcggaa gcgggttctg accagcctgt ggggttccct agaggaggca 1020

tgagctgact ctgaagggca aactgcagtt ggcctggcaa ctctggagaa gagccaccca 1080

gcataggaga cggcagtgca gaggcaccca gggcctgatg ctgggagctc agagcagctg 1140

ggggtgtgta tgcctgaaga atggtcgtag atgacactgg agggtggatt agatcaagga 1200

aagccatttg cagtgcggga gagttagtgc ctaccccaga gtctggagcc accgtatgtg 1260

atcaggttca cattttggaa agatcactgt gggaacaata tagaaagaag atgatttagg 1320

atgaggttag atttgggatc cagagatcag ttaactgtcc ccacagttag ccatattaaa 1380

gtggttaagg aaattgactg ccaatcttta gtttctttat caataaaatg aataattcct 1440

aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaa 1475

<210> SEQ ID NO 8
<211> LENGTH: 9074
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 8

ggcggcctag aagatggcgg acggcgacag cggcagcgag cgcggcggcg gcggtgggcc 60

gtgcgggttc cagcccgcgt cccgcggcgg cggcgagcaa gagacgcagg agctggcctc 120

gaagcggctg gacatccaga acaagcgctt ctacttagat gtgaagcaga acgccaaggg 180

ccgcttcctc aagatcgccg aggtgggcgc gggcggttcc aagagccgcc tcacgctgtc 240

catggcggtg gccgccgagt tccgcgactc gctgggcgac ttcatagaac actacgcgca 300

gctgggccct agcagccccg agcagctggc ggctggcgcc gaggagggcg gcgggccgcg 360

-continued

gcgcgcgctc aagagcgaat tcttggtgcg tgagaaccgc aagtactacc tggacctcaa 420

ggagaaccag cgcggccgct tcctgcgcat ccgccaaacg gtcaaccgcg gcggtggcgg 480

cttcggcgcg ggccccgggc cgggcggctt gcagagcggc cagaccatcg cgctgcctgc 540

gcagggcctc atcgagttcc gcgacgcgct ggcgaagctc atagacgact acggaggcga 600

ggacgacgag ctggcaggcg gcccgggagg cggcgccggg ggcccagggg gcggcctgta 660

tggagagctc ccggagggca cctccatcac cgtggactcc aagcgcttct tcttcgatgt 720

gggctgcaac aaatacgggg tgtttctgcg agtgagcgag gtgaagccgt cctaccgcaa 780

tgccatcacc gtacccttca aagcctgggg caagttcgga ggcgcctttt gccggtatgc 840

ggatgagatg aaagaaatcc aggaacgaca gagggataag ctttatgagc gacgtggtgg 900

gggcagcggc ggcggcgaag agtcagaggg tgaggaggtg gatgaggatt gaaacgggca 960

gcttccccta caggcctcca cccaaccacc atccccttgg ctagagaatt ccctttcctg 1020

ctcatcccca gtgagctagt ggagagggag caagggaggc cgcagaggga aaaaacaaaa 1080

cgtaacacag ttaagagaaa taatcgtaag agaacagtga cggacaactt gagaaaagca 1140

gtcaagttcc aaggaactga cagcaacctg caaagaggaa aacagcatct cctcacctgc 1200

gtaaaattgt ctcagcttct gttgtttctc aactgaggtt cgtaaaccca tcaggataat 1260

ccctggaggg aatagatcct tgcacatcca gggcaagaaa catgtccaag ttacccagac 1320

cattgataac agttgcattt aggttgcacc tgggtaatct ggcataaaag atctctctag 1380

gcctcactgt tgcggtgtct atcccttcac ctccattgaa atcagcattt tggatctagg 1440

tcttcatgga atccttgaga agagaggcct ttacaattac ccagttctga gggttcaggt 1500

tcacgaaaag aaatgcaact tgggataatc atgaacaggt taaagataag atttcaagaa 1560

gccatctaag aatacagaac caaattggat ccatttttt aaaaaaatgg ttttgcatgg 1620

aacctggacc aaggcaaatg tcttttcttc gcagaattgt tttccaggat gccagtggat 1680

tcagatagca atgcttggag tagaatccgt tactaaaata gtttcaaagt tgacaaaaaa 1740

ttttcaaaga taaaagcagt tttacattgg gggttgctga ggtaggcaca agaaaaagtc 1800

aggcataaag cacaaggcag actgtttgag tggattggtt gctgctcact aaagttgttc 1860

ccctgatctc taaatatgga ggtcattacc aagaaatgct ttggtatgaa tgagagccag 1920

atctccactg tgtgagccag tgaattatgg ctaattcggc tgttacagcc actggttggc 1980

tggattttaa accataaaac ttgaagatta cctacaaaag taacagtgtg gctataagcc 2040

tgagctttaa tggatataca tcctcacaga aaagttggaa ataaccaaaa ctgaagtctt 2100

aatttacctt cagtttaatc tgtggatttg ttcaaatact aaagatcctc aggtccagaa 2160

ttccagcatc atttattctt ttaaaatttt taagaacttg atccattgta tcagtacctc 2220

acaatcagag ttggcaaatg atggatgagt gattcaagca gtgcacccgg tggaagctga 2280

aatccatctg tgaatggaac tgaagtgaac gtgaatatgc tgactatatc ctggaagcat 2340

ttttatacca tcttgaaatt tcaacaaact ggcttttgcc agttaatcca gctgtctttc 2400

aagaataaaa gttggggttt tcaaggatcg cctcttctat attttaaatg gattttcagt 2460

agaaatgatt tttactaatc aagttaatcc caccccatca aaaggtattc ctagaaatgt 2520

catagaccta ggtaactttg aattgaatgg gagctaacgt tctttccaaa gttttcaggt 2580

attctttgtg tgacaccttc tcaaccagga ggcaagtaac cccgcctcca caatcttagt 2640

-continued

atttttttta aactgcatgc ctgccccttа tttgagctgc ctttttaatt tattgcatat 2700

cctttttatt atcttatttt ggtattattc aatctataca atctttttgt atttattggg 2760

aaatgagtaa tatacaaaaa ggttttcatg tatttgtggc tgagagggcg ggaaataatt 2820

gtgtacataa aattaggctt ttttaaaaaa atagattatg atgcagaata ttgttgatct 2880

tagattaaaa agtggaagag ccacaaacat tggtgccctt ttcagactat ttctctactc 2940

tcatcatcca cagtagaatt tttaaacaga tttttttaaa gcttttcttt taaatttttc 3000

tccgttgcaa agaatgtttc ctaaattgta tgggagcaat agtatttttg atgttttaat 3060

gacatccgta tacttgtact gtattttgta ctacaaggca gctgtttttc aataatgtcc 3120

tgctgtattt acctacgtgt tttgagtgtc tatttctttg ctgcggagaa caaattccta 3180

aatagtttta gtaaaggagc tgagaagcta gcattaggtt tgcagaaact atttaagttt 3240

caactctgag gcagcaatga aaatttaagt tgcagctatt agttgattgc tgtaactttt 3300

tcattttcaa accatgtaca attcttgtat agaccaactt gttttcttgc ttcagtggtg 3360

gttctgttgc tcagctgcag tgagccagtt caattttgca aaggtgcagt acctctcctt 3420

tttaaggggt tggtttattc ttttttcttt ttgtttggct gaattgcagt aactagcctt 3480

gcctttctat tctgtagaaa tgacagggtc ttcacaatcc ttcaccagtg gctactaagc 3540

tataattagc tgaatagaaa gaatgtggaa gtggtctgag gcatatagag tatatgccaa 3600

gaacactacc atatatggca tcagctttgg ttaccagaga aattttctta gtcattagac 3660

catgtaacag taatatatca tatgtaaatc tttagatatc aatttgaaaa tcctccaaaa 3720

aaaggagcaa agaatgcata agctatgtgt tggcaaaagt aatttatatt aaaattttga 3780

cctgcctatg taagattaag tggtaatgtc atagtggtgg gtttttaagt cttaaccaat 3840

ctctaaggtt tgtttctcct gcaagggatg gttcatggcc tctcttccca ctgcaggaag 3900

attgcagaag gctgtggatt aattgtagca tttcactgat cttcactcca gtcactaggg 3960

acaatagaaa cctgcaaaac acagattcat tcgtaaatat tattaatagc ttattaaagg 4020

aaatggtctt tgttaattcc agtcagataa tggcattgta cagacatggg aaacaaccaa 4080

ttttttgtt tttcagttgt tgctatgaat gattttgagc ctttttttt aagcttggca 4140

aacatcccag ctaatcaaaa tagtcatatt cctgagaagt aggaaactaa aacttctttt 4200

catataattg ttagaaggtt tgtttcccaa actaccatag ttacaaaggt gaaaagccaa 4260

attttaggaa cagaatcaaa agaataaaaa tctgtgaaga gatcctacta ctcttccttc 4320

tatgttttgg ttttggtttc tattgtccct atcatttcag caagtggaac agcagcaagt 4380

ttttcagtgc atatgctgca caagaacaaa atataaatct gtatggcacc aaaaatcaaa 4440

gtgaaaacca aaccaaaaac ccaaacaccc tatgtaacta tcggaggcat atacgtggta 4500

taaatgactg tagctgtgat acacacatgg ctacttgtca catcactttc cataattatt 4560

tactgcaaaa tgattgagag gcttttggtg caggcagccg ttaacctcct gcttcctttg 4620

ttacctctgg attactttgc agtaaattgc aggtctttta agagatttaa gcttcagttt 4680

tctcaaaaca aaacaattat cctgtcttat ctgaagatgc agggttgtgg gcaaaagagg 4740

ctggttataa taatgccctc atattgagtg gtctgtaaac ggctgcacac ttcaggcact 4800

gtagttgctg aagatgcttt gttaaatgtg accttgactg gctttacagg ggtgtagaat 4860

gtaatctaca caaggtgact ttgcatctat cttgctcttg aggtggatga aattgagaag 4920

-continued

ctggagtgtg taagccatgc acataagtat tcttcactgt aaattttgtt ttcattttta 4980

acccaattat ggtactttat ccaatgcaca actgatctct cagtagatat tcatttgaaa 5040

atagtgtggc cttgatcagt gagaaaggga aggagaaaag tgacttttt gcttatgtag 5100

aaatgactca tttgctgaga gtttgtcttt ctgcagcact cttggtatat agtgatcggt 5160

ctccttttg attggggaaa gttaatgttt ttgaccctgg agttaattca gttgagttat 5220

cttatatttt taggaagtat cagaattgct ctgatgaata acaaagttga ctgttttgat 5280

gtccaatctc aggttttaga atatagtggt gtaaagtccc actatttta attcttaaaa 5340

caactttaat ttcgtacacc ctaaaagtca catgcataag gcctgttcag agagcagagc 5400

ctccatcttt ttgctccttt tctactttgt acttcacttg aaaaaatatc aagtgacttt 5460

acattgtata tttccattgt aaccctgaca tttctcaaag ataaagcact ttttgatcat 5520

gaaatacatg aaatctttgt gtgatgtgga tcatagtttc tcaggctccc ttagataatt 5580

gcttatgaat attgttctaa ctctgtgtaa gaagagtaga aatctttgct aatgttagaa 5640

ggtttgtatt attgatccag aatgcatttt gctagtttcc aatggatggg agagtaaata 5700

atgctgcatt cacaatttaa taagttactt tcccttgagc cttaaggtaa ctttttcttt 5760

cctgtcaact acagcactga agttatagta agtgaatgag attatcagtt ttcagggttg 5820

gttttagagt actgtaaatc aattagctgt cttcctaaag agttacaact cccattcagt 5880

atactggata atgggtgtgt gggtggggct ggggagggcg ggagatagtt tgtagaaaag 5940

aaaaaagaaa aaaaaaaaaa cacatttacc ttaagaaaat acagacaaaa aagaataaag 6000

tatgtgtcat atgctccatt tgagattcca aatgcggccg ggcgtggtgg ctcacgcctg 6060

taatcccagc actctgggag gctgaggtgg gcgggtcagt tcaggtcagc ctggacaaca 6120

tggtgaaacc ccgtctctac taaaaatact aaaattagcc gggcatggtg gtgcatgcct 6180

gtagtcccag ctacttggga ggctgaggca ggagaatggc atgaacccag gaagcggagg 6240

ttgcagtgag ccaagatcgc gccattgcac tccagcctgg gtaacagaac aagactccgt 6300

ctcaaaaaat aaataaataa attccaaatg taaggtggaa atggcagatt ataacttcat 6360

ttcccccatt tttaacttag aaaaaacaaa gtccttaatt ttcaagtatt agccccccag 6420

aagtttcagc agatagctgg actcttgaat aatgatttct tttgatatgt ctgactttgc 6480

tatttaaagt ctcctttccc agccctcagt cttaaaactc agttttatta ctgacggatt 6540

taacagactc ttacttgaaa attgtccaat tatcagtact gaaaattaga tttggaaaat 6600

gaaaaaataa taaaatcgca tcaagaaaac ttattacctt agagattccc aattgatatc 6660

tgtacccctg atccaagtaa ggcctttatg aatgtatttt atatttccta aagaaatagg 6720

gaatatgcca tcagtgtaat cacctctatt cacagtcatt taatttgaaa atctaactca 6780

tgtctcttga attagacatt ccaattggga taggtagagt catgttaaaa caatagaata 6840

gcaactgcca tattggtagt ttcactgctc aatatgattt ctaataagag cagagcttta 6900

ttggaagcta aaactcattt ttctccatag atgattctaa cagtttaata aaaaagtata 6960

attatacttt aagtgtcaac agattttgag tctagaagca aatttcattt tccttgtgtt 7020

atatctaata tccattgaca tgggtaagaa aactaattat gtctttttag tatgatgaaa 7080

tgtgggccca gagtattgga aagaaagcaa tagccctatt cagtgtccac taggatggat 7140

attgcagaag gcatgagttt gtccttttt ccccttacc ttaaataaca cttcaatagc 7200

-continued

taacttttttc caaaatatcc atttaggcca tctcttattt cttaaagcca aatatttcgt 7260

ccccagtctt aactctgagt tgcaagatca ggctggttgg aagaatcttt atatttagaa 7320

ataggatggt gagtggtttg tttattctct tgtaaaattc ctttttcaga gacataactg 7380

cagggagagt cccaagagaa ctgaagaact ggattatatc attaaacaca tttgtgttgt 7440

aggaatttgg gagtttcatt ttgaactgat aagatattag acataaaccc ttgttccagc 7500

ccacatgtgt catccctgtg tacaaaagct ctctggattc ttttttttttt tttgagacgg 7560

agtttcgctc ttatcgtcca ggctggagtg agtgtagtgg cttgatctca gctcactgca 7620

acctttgcct cccgggctca agcgattctc ctgcctcagc ctcccaagta gctgggatta 7680

caagcatgtg ccaccatgcc cagctaattt tttgtatttt tagtagagac ggggtttcac 7740

catgttggcc aggctggtct cgaactcctg acctcaagtg atctgcccgc ctcagcttcc 7800

caaaatgctg gaattacagg catgagccat cacgcctagc ctactctctg aatttctaaa 7860

agtcagtagg ttgaccaaaa agtctagaaa ctggctttaa gtcagtatgg gacgtactta 7920

taaagagtcc atggttttgc acgtttcggt agacaagtaa atctgagtta tttttcaatg 7980

acttaccaat atttgaatag taactaagat cgtcagtgta tctggacttc tttttttgaa 8040

gttctaaaac aattatagta gggatttatt attttgggcc tccatccaga tgtttttcca 8100

agatcatttt taaaattcat ttgtcttctg tttccagata acatactttc cgttctatag 8160

gaatcttcac tgccaatcat agtatctacc agtggctttc ttagactatt cactccaaag 8220

ctgggactga tgtcctgcca gtagagaatc tacagaaata atttgaatga attaaaacca 8280

aatcttgata gcaggagaca gcttcctgat ctagatgtac aattagagtt taggttggaa 8340

attactttaa aatgtgtttt ttggggatgt cttcaatctc tgtgtaaata cccacatgct 8400

tatgcattgt aaaccaagtg tgtattcctg tgtatgaatt tgtagaactg atttctgctt 8460

caagagaagc tgcaccttta attttataag gtcccctcca cctgtaaccc tataaatgtc 8520

tgtaaataaa acactaaaat ttgtagtgat aggatcaatt tgggaatatc tgctgagaga 8580

ccaaaaagtt catttttta agtaccttgg ttaaagagta aagattattc ctcttatttt 8640

ttaaagaaga atgcacttta acaaacatag agctgcatgg gcaattcaaa caaatctgtg 8700

aagtgcagta cccattcaga aatcacactt cctgaaaacc gttcaaaagc agagtccaga 8760

cgggctgttg atctcactgc ctgtaggttg aagctcagat tctgatcaat tttgagagga 8820

gcagggctgc ttcaaaagag caatgtgaat acagtcagaa gcttcagact ggtctgtaaa 8880

aatggcgggt cccgtattta ccactaacta gcaaaactga cagaaaaact cacagagaaa 8940

aaatgtaaga atccttcctg ctggtgtgca ctccttacaa tagacttttg caaatggagt 9000

tttacagtct atatttaaaa aaaattgtat gtttgtaaca aataaagtat gcagaaaagt 9060

gaatgacaaa aaaa 9074

<210> SEQ ID NO 9
<211> LENGTH: 2843
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 9

ttattgtatg tacaaataga gaatggaaga acaaaagaat gaggaaagga agcaagaaag 60

gaaggaagaa agggtgggag ggaggaagaa cggaagggag ggaaagacaa atttcttctc 120

-continued

attctttatc cctctttgag atgcctctct gtggatcctg tcacctctac tggggagcaa 180

aaagaacaca gaagttggag ccgacggatc ccatgcctag ctgggtaatt aggcaaatct 240

cagcctctgg aaatgcagct tcttctgtaa cctacctcag agggctgttg tgaggattga 300

atggggtaac atttaaacta actggcaaaa gcaggttctc aatatattaa aattcacatg 360

aaaaagataa accaggaagg gagttggcag gaagccagga agcctcctgt gttctctgca 420

ggaggcttat ctagttcttc aggatcttgg gaacagctag tcttcctctg ctgcattctg 480

agagttaggt gatgagtcac cacctcaaga ggccatctag ttctggccga tctatctgac 540

tggatcgtgg accctaggac ccagctaact tgttctccag ccaatttaga agaatgagta 600

gcaaaaataa tatggcaagt atttcttgaa gtgccctaac cctctgtcct tattagtgta 660

ccctgtcagg agtggcctgc cttgactttc tgtcaaactg gagcagcctg agactaaggg 720

caaatgccct cttggccttg gccggccaca acttggatat gcaggtcagg cttattgtta 780

gggaagctct gagactcaat taatctgatg ccttttttatt atgagttcag gtgcagctcc 840

atttacaagt ttttctttca aaatcattca gccattaaag ctgcataatc ttggaaggtt 900

ttgaagacac atgattggta aatgtttttc cccttcacac attatctctc atgtttattc 960

atgaaaagtg aaccagtctt ataatttaaa cagtgtcttg gggtatctgg cctcctgtat 1020

ttctaagggg gaggaaagat gaggctatcc cagaagccat ttaaaaaggg gatttgaggc 1080

ctgaacgcag agaactcaat cccaggtact aatgaccttg ctttggcttg caagactaag 1140

cctgaggcct gtggcactta agagggttac tgggtaccca gacaggcaaa tgcctgttga 1200

ttggatggct aaattgtgaa gaaagcaaag cttacctgtg caatgtgtcc atgccttcat 1260

tcctactttc tgaatcaaca gtgcctgtta ttgttctgcc caactgttaa gatgttccac 1320

ctgtgcctgc tctttggcgc agtactccag gctatagacc aacatttctc aaacctggct 1380

gcacattaga gtcatctggg gatcttttaa aaacccagct gccttggctg cacatctgat 1440

cattacgtca gactcttgta gggaggagta tagacatcag cattgattaa aactcctgag 1500

ataattccaa tgtgcagtca ggattgagaa cagctggtct acaccagttc ctcttgaagt 1560

atggtctgta agtcacaggc tgcagaatct ccttggatgt ttgtgggaaa tgcagattgc 1620

taggttcacc agccctatag gcactgagtc tcaaggtgag atctggaaaa tgtgttttaa 1680

taagctcccc aggtgattcc taagcttttt gcaagtaccg ctgcttctta tgatatggcc 1740

gttcttgctt cccatcagtt atccttgtct ttcagttcct caaagatatc aagtttcatt 1800

catcctcagg accactgcat atgatagtcg ctctacttgg aagacatttt cctctttctc 1860

cttgttatta ggcatttggc agacttaaca atgaggctat tgtaacccca gaaagtcatt 1920

gcaaagtgcc tgcttagggg gttaagagta gacactgcag ggccttggaa tagtaatggt 1980

ttactattta atgagttccc tgtcactgga agtcatagag agattgctcc aggcatggtg 2040

ggagaaaagc cagtttgaca ggcggtggag ggataagagg gaagacttga gtgacgtggc 2100

acaaactggg tcagatctat taaagaggca gccctgggca gctggcaaga gatcatgtca 2160

agagacacaa tgcaggtgcc agcaggacca gaagatcagc ttcttagagc gagtgccacc 2220

tggaaggggg aaaccaaggg aaggggcaca gcaaagggca aagctgggtg gtggaggaag 2280

tacccaaaca gagggaagag ccactctggc cacatcaagt gaatactgat gatcgggaaa 2340

cattcatggg aagttatttc actctctcga gtccagtgta agtgagaaag tgtaatttaa 2400

-continued

tttaaaagag cctaatttag ccttccatca actgtagatc ttttaacgag acctttattc 2460

ttctaggaca gcatttctca aactttaatg tgcgtaggaa tcttattaaa atgcagattc 2520

tgattcagta cgtctggggt aagggctgag attcagcact tctaaccggc tcctagatgc 2580

ccccggtctg agcactacac ttagagcaag caggtttgaa cattgcgact tataacagtg 2640

gacccatggt gtagacaagt ccagaaatta ctgctgaatc aacagttttt atgttagctt 2700

ccgtatacct tcagaattca gtgtttgtgt gcattctaat tctctaaatg atgtatatta 2760

gatttccatt aacgaaaata tatgtttaag tggaactctc caattccaac ttgccaaata 2820

aaataaaaaa ttcctgtcat ttt 2843

<210> SEQ ID NO 10
<211> LENGTH: 3808
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 10

caggtgattt ttggtggggc ggggacatga aaaaaaagtt aaaatgtcct tataaagaca 60

aaatcttttt ctttcctggc tgatgatttg tcattctagt cacttcctgc cttgtgacca 120

cacacccagg cttgacaaag ctgttctgca gatcagaaag aaggggttcc tggtcataca 180

ccagtactac caaggacagc ttttttcctg caagatctgt tacctaaagc aataaaaaat 240

ggccagagga tcagtgtccg atgaggaaat gatggagctc agagaagctt ttgccaaagt 300

tgatactgat ggcaatggat acatcagctt caatgagttg aatgacttgt tcaaggctgc 360

ttgcttgcct ttgcctgggt atagagtacg agaaattaca gaaaacctga tggctacagg 420

tgatctggac caagatggaa ggatcagctt tgatgagttt atcaagattt tccatggcct 480

aaaaagcaca gatgttgcca agacctttag aaaagcaatc aataagaagg aagggatttg 540

tgcaatcggt ggtacttcag agcagtctag cgttggcacc caacactcct attcagagga 600

agaaaagtat gcctttgtca actggataaa caaagccctg gaaaatgatc ctgattgtcg 660

gcatgtcatc ccaatgaacc caaacacgaa tgatctcttt aatgctgttg gagatggcat 720

tgtcctttgt aaaatgatca acctgtcagt gccagacaca attgatgaaa gaacaatcaa 780

caaaaagaag ctaacccctt tcaccattca ggaaaatctg aacttggctc tgaactctgc 840

ctcagccatc gggtgccatg tggtcaacat aggggctgag gacctgaagg aggggaagcc 900

ttatctggtc ctgggacttc tgtggcaagt catcaagatt gggttgtttg ctgacattga 960

actcagcaga aatgaagctc tgattgctct tttgagagaa ggtgagagcc tggaggattt 1020

gatgaaactc tcccctgaag agctcttgct gaggtgggct aattaccacc tggaaaatgc 1080

aggctgcaac aaaattggca acttcagtac tgacatcaag gactcaaaag cttattacca 1140

cctgcttgag caggtggctc caaaaggaga tgaagaaggt gttcctgctg ttgttattga 1200

catgtcagga ctgcgggaga aggatgacat ccagagggca gaatgcatgc tgcagcaggc 1260

ggagaggctg ggctgccggc agtttgtcac agccacagat gttgtccgag ggaaccccaa 1320

gttgaacttg gcttttattg ccaacctctt taacagatac cctgccctgc acaaaccaga 1380

gaaccaggac attgactggg gggctcttga aggtgagacg agagaagagc ggacatttag 1440

gaactggatg aactccctgg gtgttaaccc tcgagtcaat catttgtaca gtgacttatc 1500

agatgccctg gtcatcttcc agctctatga aaagatcaaa gttcctgttg actggaacag 1560

-continued

agtaaacaaa ccgccatacc ccaaactggg aggcaatatg aagaagcttg agaattgtaa 1620

ctacgcggta gaattgggga agaatcaagc gaagttctcc ctggttggca tcggtggaca 1680

agatctcaat gaaggaaacc gcactctcac actggccttg atttggcagc taatgagaag 1740

gtatacactg aatatcctcg aagaaattgg tggtggccag aaggtcaatg atgacattat 1800

tgtcaactgg gtgaatgaaa cattgaggga agcaaagaaa agttcatcca tctctagttt 1860

caaggacccg aagattagta caagtctgcc tgttctggac ctcatcgatg ccatccaacc 1920

aggttccatt aactatgacc ttctgaagac agaaaatctg aatgatgatg agaaactcaa 1980

caatgcaaaa tatgccatct ctatggcccg aaaaattgga gcaagagtgt atgccctgcc 2040

agaagacctg gttgaagtga accccaaaat ggtcatgacc gtgtttgcct gcctcatggg 2100

gaaaggaatg aagagggtgt gaggcccaat ggggctgggt gggaggcggt gcactcactc 2160

ctgactgccc ggcacagatg ctccagggat gattcaagcc attccaaagt tcaacttggt 2220

gacactctat aagattccaa aaagcacata ttagtgcagc caagtagcct ctcctgtatt 2280

taacaaaaag tgcttcattc tttgcaggag gcccaacctc ctatatatag gtttctattc 2340

ttgatttatt tgcttcttcg aaaatctaga ggaaaagaaa gaagttattt tccaggtacc 2400

cttctcgctt ttgccattag ccaaggatag aagctgcagt ggtattaatt ttgatataat 2460

ctttcaaacc agcttcatgt ggcttccctt ttctttgttc aagatgaggg ccaggagggg 2520

aaacatcaca cctgccctaa accctgttcc tggaggtcag catttgatct gttgcaagcc 2580

cctctttctg tcccctcttc ctaccctgcc tcccttgact ttgctcctca cacttttgga 2640

accatgcctt ccgggggggc ccatctcttc tggccgtcct tgtctctggg ccacttggag 2700

tgtgtgataa atcagtcaag ctgttgaagt ctcaggagtc tctggtagcc tgcagaagta 2760

agcctcatca tcagagcctt tcctcaaaac tggagtccca aatgtcatca ggttttgttt 2820

tttttcagcc actaagaacc cctctgcttt taactctaga atttgggctt ggaccagatc 2880

taacatcttg aatactctgc cctctagagc cttcagcctt aatggaaggt tggatccaag 2940

gagggtgtaa tggagcatca agccactcgg cggcagcatg gagctataac taagcatcct 3000

ttagggttct gcctctccag gcatttagcc ctcacattag atctagttac tgtggtatgg 3060

ctaatacctg tcaacatttg gaggcaatcc taccttgctt ttgcttctag agcttagcat 3120

atctgattgt tgtcaggcca tattatcaat gtttactttt ttggtactat aaaagctttc 3180

tgccacccct aaactccagg ggggacaata tgtgccaatc aatagcaccc ctactcacat 3240

acacacacac ctagccagct gtcaagggca gaatgaatct atgctggata agaaatggtg 3300

gaactgcgtt atgaagagct aatttactgg acaaagaatt ccaaagcaaa accagaacag 3360

tatgaatttg agcaggtctc ataggttgag caatttcccc ctaaaccaac tgaaggctaa 3420

aaagcaacag gccattgtga accaatgcaa gacgccctct atcatggtga aaagctccat 3480

caatgaggta tcttctttag tggtggtatg taatggaact tagccatttt tcaaagcaat 3540

tgaaatgcat tgctctggat ctgttccttg gcagtggact cagaaagcca acatgtggct 3600

cctcccagcc cataaccagt atttttgctg cttctgaata caaattggtt ggttttgact 3660

tcagattgaa cttactgtag cctcagatga tttccccccc ccgcctccag gaagaaagaa 3720

tgttactgcc ttaataaaaa atgaaaagag aatgatgctc aaaatctttc caaataaaat 3780

gttccctata aaaaaaaaaa aaaaaaaa 3808

-continued

<210> SEQ ID NO 11
<211> LENGTH: 2409
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 11

tattctggag ttctaatata ttttcatggt ttttaaatgt ttagttcttt aatcaagttg 60

gaatttttaa aatatatatg aagtaggatg ctaatttaaa acttttcctc caaatggata 120

tccatccagt tgtctcaaca tcatgtatag aaaaaacaaa ttgaaaatgc tatcttatca 180

tattatttac taaatttaca tgcatataca tatacataca taacatatat ctgagtttga 240

gtctgaattc actactgttt caattgctgt agttggtcaa ttctagtatg acagagtttt 300

aattgctaca gttttggaat gattatgggg ttacttttga gacagggtct ggctatcttg 360

cccagactgg tctcaaactt accggctcaa gcgatcctcc catctcagcc tcttaaagtg 420

ttgggattat aggtatgaac cactgcacct ggccctggat tgcattttga ctgtggccat 480

agaagtttcc actcagatct tcaagagaac ctactctggg gactacaagt gactgacagc 540

cagctgccac ccctttggaa ctaccatgca ttcatgtcaa tgccacatat ctctggagtc 600

gcccccagct agtggctgag cacagcagga gtactaaact tgttctttct tgcccagtgc 660

agaactcctc taacgggcaa ctttggcttg gggaacttct catcagcctg gatgagacta 720

tcttagaact acagtgtagc cgaggctctt tctacccaat atttccttct cttcttttta 780

cagaagtcag acctgcacca cagtctgaag tctctctgcc aatccctgag ccttcccact 840

ttatccttca caaacatttc ccccaataaa tatcttgcat gtctaatccc ttcttggcat 900

ctgcttctct gaggaccaga acagacaggt ggtttaccaa gtggtgcaag aaaacagata 960

gtaagatggg ttttgggcaa tggcttactt ggtccctggc acaaagcagc catctgagtc 1020

ctgatttagc cagcagtgac ctaggaaaac atcccagaga agggaaatcc tcttgctcgt 1080

gcaatttagg catttaaaag atatggggta tacgatgcct acaaagaaag cagagctggc 1140

caactgttac taagacaagt aatgccttgc agatggataa tgagaaaccg agagcactta 1200

aacagttaaa ggttaagtgt gggagccggg ggcttctctg gagcttaaca aacttttatt 1260

tcctgcagcg gaagagcaga aatgggattg gatggtcaca gagctctaaa gatatttgaa 1320

tgctcagcct agggagatcc agtatgtagt cagggccctg gttggggaaa ccagcctgaa 1380

gctttgaaac atggaatggg tgcccctgag gattttggct ctgtagatac ctctggacac 1440

tcaagagctt acagaactgg ctcatacttc ctagtaagag caagcacttc cctcattaca 1500

gaaagacctc ttccataatt gttctcagtt tttctcctgg ctgccaagcc atggtaactg 1560

gggttaacat ctcagcataa tctgcctagg aactgctggc cctgataaga aaggaaagat 1620

tacacaccga aggaattgaa ataagttacc atgactatca ggagccagag aagtatactt 1680

aggattggat tttaagggtg ttcaattaag ggggctggaa tttaagacaa cataagcaag 1740

acttcatgga cttggggtac ttttttggga cagatttaac ctcagcaagt agtcccgagg 1800

ttagcttttc aaagcctgga aaaggcgatg gctttgaaaa agtggaaatg aattcacacc 1860

atggcagctt ttaacttgag gagaagatag ctatttattt tggcatattt atgtggttac 1920

tgctcactta ctagtcgcaa ttgttttgta accacttttt atttatttt tgagatagtc 1980

tcattctgtt ttctagactg tgctacagcc tgggcgacac agtgagactc cgtctcaaaa 2040

-continued

aaaagaaaaa aaaaaaaata tcatattgta ttggccatcc cttgtttcat attcgttaat 2100

aaactttgct actttgcttt ctaatttcgg tgaagaacaa gcacaagatt aaggagtaag 2160

actatgtgaa actgggttat aaattgaaag gctgttggct gttgtatctt gtaaacaatt 2220

caacgcaatg ccaagtgaaa atgaagttgt tagtccctct atctagaatt ctttcccaaa 2280

tagctgcata tctttttcct ctgcttcatt taagtctgtg ctgaaatgtt atctactcag 2340

agaagacatt ctggatcatt tcatattaaa tagtaacttt tatcactcca aaaaaaaaaa 2400

aaaaaaaaa 2409

<210> SEQ ID NO 12
<211> LENGTH: 1338
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 12

gcccctgcta ggacagcccg tgcgagcctg ctggaggagg aagagaaagg cagagagagt 60

cgggttacaa gatggcggat ctgtagtagt taccgcggcg gcgggagagc aagcgagccc 120

tgggggggcaa agagacggga gagtgggtgt atgcgcgggt gaagtgagag gtaacggggc 180

tccgggcgga gaggcctcag tggctcttgt caccccttct cgcggctgaa cctttggagc 240

catggtgaat tcgggcctct ccgaagccgc cgccgccgcc accgccacta ctgcctttac 300

cgtctcctaa gagtgaggag cgcggacgag gtaagcgagg aggcggcggc tagagcggtg 360

gagacagcag ccaccatgtc ggatacgcgg cggcgagtga aggtctatac cctgaacgaa 420

gaccggcaat gggacgaccg aggcaccggg cacgtctcct ccacttacgt ggaggagctc 480

aaggggatgt cgctgctggt tcgggcagag tccgacggat cactactctt ggaatcaaag 540

ataaatccaa atactgcata tcagaaacaa caggcaagta tttcagactg aaaaaaagtt 600

tttgtactgg tgataattat cattttgga ttgagccact gtcggtttat tctaagatgt 660

atttattagt attatttaac tgtagttagc caagtctctt ctataccttg tacatgaaac 720

cttttattct gagtcacgtt taaaggaaat gacatagaac aaaattcaaa gggatactct 780

taaagcaagg aaagcctgtt tgtggtgtga aattaaatcc tttgactcaa gtaaaggtag 840

aggttttgaa gtagaagcat taaatcctgc tctttaattg ttgtgtattt ttaatatagt 900

aatcaatgga tactttataa aaatgtacat taccaagtat gtaacatctt ctgttaatat 960

tgtcttattt gggaaagaaa atgtgtattg tattttagca cgcctgatta atctgaaaat 1020

ggtagcagtg attcaatctt ttttattctg tgaatgtttt tgtttttatt atgaaatata 1080

ttaaaattga aaagtacaga gaataatcgg tgtctactgt aacaccacca gatttaacaa 1140

tgttaatgtt acgccatatt tgtttcaaat atttttgtaa tattgaacat tatggataga 1200

gttaaagctt gtttgtatcc atcccgttgt ttacattctc catcccctac ataggtaacc 1260

actattctga agttgatgtg tattctttgt gtacatgctt ttataccttt tctgcatatg 1320

tatgtatcca taaataat 1338

<210> SEQ ID NO 13
<211> LENGTH: 719
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 13

-continued

gacaggcttg gcaggtacta agccagatgg aggcatgacc ctgggagggg gtggagtgca 60

aagaggtgca gagcacccca taaaataagg ggttgggcat caccatgccc cctgcaggct 120

ttgccaccca tcagcaccag caagtcgccc cacccagagc tctgcaaaga acaccgccag 180

ctctttcctt ccgggtgctc ttgctggggg gtcaggctgg tcccctctga ctggcactgc 240

ccacgctggg acatgggagg agggccatct gggcactgca ggagcagatc ctgggtgcac 300

atggggagac aggtgtgctg gcaggaacag caggcggggc tgagctcagg aggcgcctca 360

cagcgggtgc ctccagcctg gtctgagcac gggatgaaga accaggcacc gctcatcgtg 420

catctccaca aggcactgtg tgacttcatg caagccgctc aacctctctg gccccacgtc 480

cttgtccaca ggggcagagt ggactgaatg gaggctgagg tccctctcag cttgggtggt 540

gtggcagagg gtgaggagga gcccatgggc cctggatcac cagcctccct ccacccagaa 600

tgccccgtcc actccactgc ccaccaccct ctgtgcaatt gacaaacgtc ctggtgttaa 660

cgcctaagca taaagagtcg gctttgctgt aaaaaaaaaa aaaaaaaaaa aaaaaaaaa 719

<210> SEQ ID NO 14
<211> LENGTH: 2985
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 14

tcgggcacgg cgtcctccct ccgcagcagc cgagccggac ctgcctcccc gggcgtgctc 60

cgccggcccc gccgccggcc cgcagcgaca gacaggcgct ccccgcagct ccgcacggga 120

cccaggccgc cggaccccag cgccggacca ccctctgtcc gccccgagga gtttgccgcc 180

tgccggagca cctgcgcaca gatggagctg gaccaccgga ccagcggcgg gctccacgcc 240

taccccgggc cgcggggcgg gcaggtggcc aagcccaacg tgatcctgca gatcgggaag 300

tgccgggccg agatgctgga gcacgtgcgg cggacgcacc ggcacctgct ggccgaggtg 360

tccaagcagg tggagcgcga gctgaagggg ctgcaccggt cggtcgggaa gctggagagc 420

aacctggacg gctacgtgcc cacgagcgac tcgcagcgct ggaagaagtc catcaaggcc 480

tgcctgtgcc gctgccagga gaccatcgcc aacctggagc gctgggtcaa gcgcgagatg 540

cacgtgtggc gcgaggtgtt ctaccgcctg gagcgctggg ccgaccgcct ggagtccacg 600

ggcggcaagt acccggtggg cagcgagtca gcccgccaca ccgtttccgt gggcgtgggg 660

ggtcccgaga gctactgcca cgaggcagac ggctacgact acaccgtcag cccctacgcc 720

atcaccccgc ccccagccgc tggcgagctg cccgggcagg agcccgccga ggcccagcag 780

taccagccgt gggtccccgg cgaggacggg cagcccagcc ccggcgtgga cacgcagatc 840

ttcgaggacc ctcgagagtt cctgagccac ctagaggagt acttgcggca ggtgggcggc 900

tctgaggagt actggctgtc ccagatccag aatcacatga acgggccggc caagaagtgg 960

tgggagttca agcagggctc cgtgaagaac tgggtggagt tcaagaagga gttcctgcag 1020

tacagcgagg gcacgctgtc ccgagaggcc atccagcgcg agctggacct gccgcagaag 1080

cagggcgagc cgctggacca gttcctgtgg cgcaagcggg acctgtacca gacgctctac 1140

gtggacgcgg acgaggagga gatcatccag tacgtggtgg gcaccctgca gcccaagctc 1200

aagcgtttcc tgcgccaccc cctgcccaag accctggagc agctcatcca gaggggcatg 1260

gaggtgcagg atgacctgga gcaggcggcc gagccggccg gcccccacct cccggtggag 1320

-continued

gatgaggcgg agaccctcac gcccgccccc aacagcgagt ccgtggccag tgaccggacc 1380

cagcccgagt agagggcatc ccggagcccc cagcctgccc actacatcca gcctgtggct 1440

ttgcccacca ggacttttga gctggggctg actcctgcag gggaagccct ggtccagctg 1500

ggtgcccccc cgagctccgg gcggactcgc acacactcgt gtcatccaga tgtgagcacc 1560

gcacccagcg gcaaagagcc ctccccccctg cagggctcca cccatcaccc tccctccgtc 1620

tgtctttccg gcctggaccc caccctccac actctcaggc catcacagaa caccccagct 1680

tcctcattct gctacaacac ccaggccctc tggacatcca gaaaaccaag tgtccggatg 1740

gcaggggcca gcggccacca agctcatggg acacccagag cagaagctag ggcagagcca 1800

atgctgaggg agcctcgact tccggcgccg ccgccctctc ccggcatccg cagagccagc 1860

tgacgccctc cctgcctccc agggcagctg gccagcctcg ggcagcgcgg ccccctcctc 1920

ccaggggaga gtagaagtcg cacacgcagc agagcagacc tgatgtcccg gtgcttcctg 1980

gcccctcagc tccagtgatt cacgcccgcc tggagaagaa tcagagctca gctcatgact 2040

cacccatggc aggcggaggg tcccagaggg gctgagtcct caaatccggc tgaggcagca 2100

gctggcacca tcagagccag gagagtgaca acaggtctca aggttcccac aaagtctttg 2160

ctgctgtgct gggcaccacc cacccctcac cttgcaggct gcctgcgtgg gaggcgaagt 2220

cccaggacag cccagagggg ggctacagag aggagtcggc tgcagcagag ggcaggagcc 2280

ccagcttagc cctgagcgcc agcgcgagga ccagggcctg ccactaagcc cgccccgctg 2340

gccgccagct gcccgtcccc agagccactg cagcaggagt cgggccctgc ctccctccca 2400

gcagggaaac cccgcccgct gccaggccat cctctctgcc agaggctttc atgagcccca 2460

aggctggggc cacagctcct acccctgccc agcagccctg agctcagctg caggaaggac 2520

atcccagaag ccatggctcc tggggcgctt ccaggcattc tgccctgccc cgacaccaga 2580

accctggtgc tggtgggcca ctagcgtctg cagcctaagc aggtgctggc tcagggttca 2640

tcgttctgcc ttgtccactg ggggaccagc cctgcagacc actctgacaa gtcttcagcc 2700

cacaccctgc cagccccaca gattttattt ttgcacataa gccataacca atcctcaagg 2760

ctggcacagg ctttggggaa gccctggagc ctgtgaagac cctggaaacc tcatgaggct 2820

gtggccaacc cctgcccctt gccccacaca gaccaggcct taaatgtcgg tccaggccct 2880

gtgcacctta ccccagagac agactctttt tgtaagattt tgttaataaa acactgaaac 2940

ttcaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaa 2985

<210> SEQ ID NO 15
<211> LENGTH: 848
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(1)
<223> OTHER INFORMATION: n is a, c, g, or t
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (33)..(33)
<223> OTHER INFORMATION: n is a, c, g, or t
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (39)..(39)
<223> OTHER INFORMATION: n is a, c, g, or t
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (92)..(92)

-continued

<223> OTHER INFORMATION: n is a, c, g, or t
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (207)..(207)
<223> OTHER INFORMATION: n is a, c, g, or t
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (217)..(217)
<223> OTHER INFORMATION: n is a, c, g, or t
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (228)..(228)
<223> OTHER INFORMATION: n is a, c, g, or t
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (267)..(267)
<223> OTHER INFORMATION: n is a, c, g, or t
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (314)..(314)
<223> OTHER INFORMATION: n is a, c, g, or t
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (420)..(420)
<223> OTHER INFORMATION: n is a, c, g, or t
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (449)..(449)
<223> OTHER INFORMATION: n is a, c, g, or t
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (460)..(460)
<223> OTHER INFORMATION: n is a, c, g, or t
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (477)..(477)
<223> OTHER INFORMATION: n is a, c, g, or t
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (490)..(490)
<223> OTHER INFORMATION: n is a, c, g, or t
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (499)..(499)
<223> OTHER INFORMATION: n is a, c, g, or t
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (540)..(540)
<223> OTHER INFORMATION: n is a, c, g, or t

<400> SEQUENCE: 15

ngaggaacaa aagtcgtttt gtttggtgat gcnccgggna aggcagtgag ctgtttgctt 60

ttttacagcg aaggtgagag gcaagttatc tntatttgtg gacttcaaag aggcgtgttc 120

ggaacaagtc ccaaagcttt ggcatgatgt ttattagcaa agttcagtga attgcaaatt 180

aatacccaca gtggcattta tgaatgnggg ggggggnctt ccctcttntt gggtatgttg 240

tacttgtttg tttttggagg ttcccanttt agttccagcc cactgcctgg gttggaagtc 300

cgagaggaaa cgangagttc tccgggaggc tgacttatac ttcctcacgg atgcaggaga 360

cggggtaagc ccttcccgcc tccctccgcg ggcaggaggc tatgctaacc aagcagccan 420

cagaagtgca ggagaccacc gtggagtgnc ctgaaatcan cacactccag tgaaganatt 480

tgctagaggn ggggtgggng acttttcaag acaatggatt gtcaggtgtt atgcaaatan 540

agaatggaga ggagggttcc tggcttctgt tgtggcgtct tttctatttg acttcatggt 600

tgtaagtatt tccagatggt gaatcagaca ccagacgtaa caatatttcc atatttgggt 660

atagcagtag cctgcgtttt taacattttt ctgcctctgt tatccaacca caaagcattc 720

-continued

ttgacagctt caaatgttgt taaatataga tttaacttct cttcccagag caggaaattc 780

tttggaattc cttgtttttc acgcaatctg tccatcatga tttaaaataa aagcacagtg 840

gatcatcc 848

<210> SEQ ID NO 16
<211> LENGTH: 2295
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 16

agaccgcgag cgagagcgcc cccgagcagc gcccgcgccc tccgcgcctt ctccgccggg 60

acctcgagcg aaagacgccc gcccgccgcc cagccctcgc ctccctgccc accgggccca 120

ccgcgccgcc accccgaccc cgctgcgcac ggcctgtccg ctgcacacca gcttgttggc 180

gtcttcgtcg ccgcgctcgc cccgggctac tcctgcgcgc cacaatgagc tcccgcatcg 240

ccagggcgct cgccttagtc gtcacccttc tccacttgac caggctggcg ctctccacct 300

gccccgctgc ctgccactgc cccctggagg cgcccaagtg cgcgccggga gtcgggctgg 360

tccgggacgg ctgcggctgc tgtaaggtct gcgccaagca gctcaacgag gactgcagca 420

aaacgcagcc ctgcgaccac accaaggggc tggaatgcaa cttcggcgcc agctccaccg 480

ctctgaaggg gatctgcaga gctcagtcag agggcagacc ctgtgaatat aactccagaa 540

tctaccaaaa cggggaaagt ttccagccca actgtaaaca tcagtgcaca tgtattgatg 600

gcgccgtggg ctgcattcct ctgtgtcccc aagaactatc tctccccaac ttgggctgtc 660

ccaaccctcg gctggtcaaa gttaccgggc agtgctgcga ggagtgggtc tgtgacgagg 720

atagtatcaa ggaccccatg gaggaccagg acggcctcct tggcaaggag ctgggattcg 780

atgcctccga ggtggagttg acgagaaaca atgaattgat tgcagttgga aaaggcagct 840

cactgaagcg gctccctgtt tttggaatgg agcctcgcat cctatacaac cctttacaag 900

gccagaaatg tattgttcaa acaacttcat ggtcccagtg ctcaaagacc tgtggaactg 960

gtatctccac acgagttacc aatgacaacc ctgagtgccg ccttgtgaaa gaaacccgga 1020

tttgtgaggt gcggccttgt ggacagccag tgtacagcag cctgaaaaag ggcaagaaat 1080

gcagcaagac caagaaatcc cccgaaccag tcaggtttac ttacgctgga tgtttgagtg 1140

tgaagaaata ccggcccaag tactgcggtt cctgcgtgga cggccgatgc tgcacgcccc 1200

agctgaccag gactgtgaag atgcggttcc gctgcgaaga tggggagaca ttttccaaga 1260

acgtcatgat gatccagtcc tgcaaatgca actacaactg cccgcatgcc aatgaagcag 1320

cgtttccctt ctacaggctg ttcaatgaca ttcacaaatt tagggactaa atgctacctg 1380

ggtttccagg gcacacctag acaaacaagg gagaagagtg tcagaatcag aatcatggag 1440

aaaatgggcg ggggtggtgt gggtgatggg actcattgta gaaaggaagc cttgctcatt 1500

cttgaggagc attaaggtat ttcgaaactg ccaagggtgc tggtgcggat ggacactaat 1560

gcagccacga ttggagaata ctttgcttca tagtattgga gcacatgtta ctgcttcatt 1620

ttggagcttg tggagttgat gactttctgt tttctgtttg taaattattt gctaagcata 1680

ttttctctag gctttttttcc ttttggggtt ctacagtcgt aaaagagata ataagattag 1740

ttggacagtt taaagctttt attcgtcctt tgacaaaagt aaatgggagg gcattccatc 1800

ccttcctgaa gggggacact ccatgagtgt ctgtgagagg cagctatctg cactctaaac 1860

-continued

tgcaaacaga aatcaggtgt tttaagactg aatgttttat ttatcaaaat gtagcttttg 1920

gggagggagg ggaaatgtaa tactggaata atttgtaaat gattttaatt ttatattcag 1980

tgaaaagatt ttatttatgg aattaaccat ttaataaaga aatatttacc taatatctga 2040

gtgtatgcca ttcggtattt ttagaggtgc tccaaagtca ttaggaacaa cctagctcac 2100

gtactcaatt attcaaacag gacttattgg gatacagcag tgaattaagc tattaaaata 2160

agataatgat tgcttttata ccttcagtag agaaaagtct ttgcatataa agtaatgttt 2220

aaaaaacatg tattgaacac gacattgtat gaagcacaat aaagattctg aagctaaatt 2280

tgtgatttaa gaaaa 2295

<210> SEQ ID NO 17
<211> LENGTH: 340
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 17

atgggcggag ggaagctcat cagtggggcc acgagctgag tgcgtcctgt cactccactc 60

ccatgtccct tgggaaggtc tgagactagg gccagaggcg gccctaacag ggctctccct 120

gagcttcagg gaggtgagtt cccagagaac ggggctccgc gcgaggtcag actgggcagg 180

agatgccgtg gaccccgccc ttcggggagg ggcccggcgg atgcctcctt tgccggagct 240

tggaacagac tcacggccag cgaagtgagt tcaatggctg aggtgaggta ccccgcaggg 300

gacctcataa cccaattcag accactctcc tccgcccatt 340

<210> SEQ ID NO 18
<211> LENGTH: 2037
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 18

catcttctaa attgagcctc cggtcatact agttttgtgc ttggaacctt gcttcaagaa 60

gatccctaag ctgtagaaca ttttaacgtt gatgccacaa cgcagattga tgccttgtag 120

atggagcttg cagatggagc cccgtgacct ctcacctacc cacctgtttg cctgccttct 180

tgtgcgtttc tcggagaagt tcttagcctg atgaaataac ttggggcgtt gaagagctgt 240

ttaattttaa atgccttaga ctggggatat attagaggaa gcagattgtc aaattaaggg 300

tgtcattgtg ttgtgctaaa cgctgggagg gtacaagttg gccattccta aatctgtgtg 360

tgagaaatgg caggtctagt ttgggcattg tgattgcatt gcagattact aggagaaggg 420

aatggtgggt acaccggtag tgctcttttg ttcttgcttc gttttttaa acttgaactt 480

tacttcgtta gatttcataa tactttcttg gaattctctg gggctttggg gaatttagtg 540

cgtgggtgag ccaagaaaat actaattaat aatagtaagt tgttagtgtt ggttaagttg 600

ttgcttggaa gtgagaagtt gcttagaaac tttccaaagt gcttagaact ttaagtgcaa 660

acagacaaac taacaaacaa aaattgtttt gctttgctac aaggtgggga agactgaaga 720

agtgttaact gaaaacaggt gacacagagt caccagtttt ccgagaacca aagggagggg 780

tgtgtgatgc catctcacag gcaggggaaa tgtctttacc agcttcctcc tggtggccaa 840

gacagcctgt ttcagagggt tgttttgttt ggggtgtggg tgttatcaag tgaattagtc 900

acttgaaaga tgggcgtcag acttgcatac gcagcagatc agcatccttc gctgcccctt 960

-continued

agcaacttag gtggttgatt tgaaactgtg aaggtgtgat tttttcagga gctggaagtc 1020

ttagaaaagc cttgtaaatg cctatattgt gggcttttaa cgtatttaag ggaccactta 1080

agacgagatt agatgggctc ttctggattt gttcctcatt tgtcacaggt gtcttgtgat 1140

tgaaaatcat gagcgaagtg aaattgcatt gaatttcaag ggaatttagt atgtaaatcg 1200

tgccttagaa acacatctgt tgtcttttct gtgtttggtc gatattaata atggcaaaat 1260

ttttgcctat ctagtatctt caaattgtag tctttgtaac aaccaaataa ccttttgtgg 1320

tcactgtaaa attaatattt ggtagacaga atccatgtac ctttgctaag gttagaatga 1380

ataatttatt gtatttttaa tttgaatgtt tgtgcttttt aaatgagcca agactagagg 1440

ggaaactatc acctaaaatc agtttggaaa acaagaccta aaaagggaag gggatgggga 1500

ttgtggggag agagtgggcg aggtgccttt actacatgtg tgatctgaaa accctgcttg 1560

gttctgagct gcgtctattg aattggtaaa gtaataccaa tggcttttta tcatttcctt 1620

cttcccttta agtttcactt gaaattttaa aaatcatggt tatttttatc gttgggatct 1680

ttctgtcttc tgggttccat tttttaaatg tttaaaaata tgttgacatg gtagttcagt 1740

tcttaaccaa tgacttgggg atgatgcaaa caattactgt cgttgggatt tagagtgtat 1800

tagtcacgca tgtatgggga agtagtctcg ggtatgctgt tgtgaaattg aaactgtaaa 1860

agtagatggt tgaaagtact ggtatgttgc tctgtatggt aagaactaat tctgttacgt 1920

catgtacata attactaatc acttttcttc ccctttacag cacaaataaa gtttgagttc 1980

taaactcaaa aaaaaaaaaa aaaaaaaaaa aaaaaaagaa aaaaaaaaaa aaaaaaa 2037

<210> SEQ ID NO 19
<211> LENGTH: 1338
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 19

cggatcgcct ctcggatgcc gcttgcctgg aagctgcgtt aggagcgagc ggcggcggtg 60

gcggcggtgg cggcggcggc ggcagctcgg gagtgctatg accggcaaac tcgccgagaa 120

gctgccggtg accatgagca gtttgctaaa ccaactgcct gacaatctgt accccgagga 180

gatccccagc gcgctcaacc tcttctccgg cagcagcgac tcggtagtcc attacaatca 240

gatggctaca gagaatgtaa tggacatcgg tctgaccaac gagaagccca acccggaact 300

ctcttactcc ggctccttcc agccagcccc cggcaacaag accgtgacct acttgggaaa 360

gttcgccttc gactcccctt ccaactggtg ccaggacaac atcattagcc tcatgagcgc 420

cggcatcttg ggggtgcccc cggcttcagg ggcgctcagc acgcagacgt ccacggccag 480

catggtgcag ccaccgcagg gtgacgtgga ggccatgtat cccgcgctac ccccctactc 540

caactgcggc gacctctact cagagcccgt gtctttccac gacccccagg gcaatcccgg 600

gctcgcctat tccccccagg attaccaatc ggccaagccg gcgttggaca gcaatctctt 660

ccccatgatt cctgactaca acctctacca ccaccccaac gacatgggct ccattccgga 720

gcacaagccc ttccagggca tggaccccat ccgggtcaac ccgcccccta ttacccctct 780

ggagaccatc aaggcattca aagacaagca gatccacccg ggctttggca gcctgcccca 840

gccgccgctc accctcaagc ccatccggcc ccgcaagtac cccaaccggc ctagcaagac 900

accgctccac gaacggcccc acgcgtgccc ggccgagggc tgcgaccgcc gtttcagccg 960

-continued

ttcggacgag ctgacccggc acctgcgcat ccacacgggc cacaagccct tccagtgccg 1020

gatctgcatg cggagcttca gccgcagcga ccacctcacc actcacatcc gcactcatac 1080

gggcgagaag ccctttgcct gcgagttctg cgggcgcaag tttgcgcgca gcgacgagcg 1140

caagcgccac gccaagatcc acctcaagca aaaggagaag aaggcggaga agggcggtgc 1200

accctctgca tcctcggcgc cccccgtgtc gctggccccc gtggtcacca cctgcgcctg 1260

aggatcgggc ccccagatcc ccacttttcc cctccagtgc ctcccggctg ctagcctgaa 1320

agcagcggga aagccagc 1338

<210> SEQ ID NO 20
<211> LENGTH: 1635
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 20

tgctcttccc gcgctgtccc gccgctcctg acccccttcag actgtcccga gaagtctagg 60

aaggcaccgg agaccctcgg cacaaggcac tgaacctgga gcgcctgagg ggaactgacg 120

cttgcttggc acctttggga agactgcgaa ccagctcagg tcccagggga gcgtgtccgt 180

gtctttttcg gccgcattga tgcacccgcc caagtcattt cctcagtctt ctgcttccta 240

gaagcaaaac aaggaataac attatgatac ctttaaaaag ctgaggcgtg aaactgctcc 300

tgcaacttgg aagcgcaggg cacttgaact ttattcccta aatgctcgaa aaagaaatct 360

ttctggtggc agcttgctac tcaagatgag tgactctgac cacaaccggc tgaaattaac 420

ttgaggccct agatcatttg ctatgacact aattccacgt tgttcctcaa ttatgtctcc 480

attacctcaa aggctgaagt cgtcctggct gcacggagca gcccagtctt gtctctgtaa 540

ataccgcaga tttgaaaaat agtcctattg gaacccgtca gagggatgta atacaaggga 600

ctgtgatggc tcttgaacgg agccagagac tcttcctgaa ggactggact acctccccgc 660

ctaggaaggt tgggtaaccc aaagttgcca catttctggg ctgcacacca aggagaggcc 720

ccggccactc cacacaggaa atgcacgtct gggcacctgg agccgcagga cctgagccac 780

taactgtccc tggatccctt ctgcttcaga ggcaaggaag actgccggag tcattgcttg 840

cctgctggct tcaccagcgc tttctgtttt tccagttctc ctcaccactc ttcctggcgc 900

ctttctctgc ctcccgcacc cctgcctcct ctaaaaccca cagcatcatc cctcagaaca 960

gctggaagcg gacagctgat ccttctcctt gctggtcccc cagcccagaa cacagcctgg 1020

gacatacgtg cggagccaca gtgtctgttg gtctcacctg agagctgctt gtcactccca 1080

caccctcttt caaggagaag ctgtcagcat tgtccttctg cagcattcaa caggggccag 1140

gcattgggtt agggctgaaa acactgatgg agagccatca tctttgcctt ctaaagggct 1200

tggactgtga taagaaggtg gacacttgaa gtggctgcag gaccggcatt ccagtgctgg 1260

aggggtgttt agtcctccat cagctcagag aagtaaaggc acatttattg gcagggaaaa 1320

cccaaggagg gcttctcagg agagacacac atttgactct gggggaaagt ggtccatagt 1380

accatgtaca aaggtttaat ttgctaatac attttttcaa atgccatttt gatattttta 1440

actcataaag agacaacact ttccaaacct gctgccagca atgaatctac attgtgtctg 1500

taaagccaga acaggtattt gatggacagc tgaaacaatc tttataatct aagatgctaa 1560

tcaatgatat taacatgttg ctcaattatt taaggttagt agagtcacag tcgaatctaa 1620

-continued

gatttatatt aaaac 1635

<210> SEQ ID NO 21
<211> LENGTH: 8110
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 21

gatcagagtg ggccactgcc agccaacggc ccccggggct caggcgggga gcagctctgt 60

ggtgtgggat tgaggcgttt tccaagagtg ggttttcacg tttctaagat ttcccaagca 120

gacagcccgt gctgctccga tttctcgaac aaaaaagcaa aacgtgtggc tgtcttggga 180

gcaagtcgca ggactgcaag cagttggggg agaaagtccg ccattttgcc acttctcaac 240

cgtccctgca aggctggggc tcagttgcgt aatggaaagt aaagccctga actatcacac 300

tttaatcttc cttcaaaagg tggtaaacta tacctactgt ccctcaagag aacacaagaa 360

gtgctttaag aggcggcgga aggtgatcga attccggtga tgcgagttgt tctccgtcta 420

taaatacgcc tcgcccgagc tgtgcggtag gcattgaggc agccagcgca ggggcttctg 480

ctgaggggggc aggcggagct tgaggaaacc gcagataagt ttttttctct ttgaaagata 540

gagattaata caactactta aaaaatatag tcaataggtt actaagatat tgcttagcgt 600

taagttttta acgtaatttt aatagcttaa gattttaaga gaaaatatga agacttagaa 660

gagtagcatg aggaaggaaa agataaaagg tttctaaaac atgacggagg ttgagatgaa 720

gcttcttcat ggagtaaaaa atgtatttaa aagaaaattg agagaaagga ctacagagcc 780

ccgaattaat accaatagaa gggcaatgct tttagattaa aatgaaggtg acttaaacag 840

cttaaagttt agtttaaaag ttgtaggtga ttaaaataat ttgaaggcga tcttttaaaa 900

agagattaaa ccgaaggtga ttaaaagacc ttgaaatcca tgacgcaggg agaattgcgt 960

catttaaagc ctagttaacg catttactaa acgcagacga aaatggaaag attaattggg 1020

agtggtagga tgaaacaatt tggagaagat agaagtttga agtggaaaac tggaagacag 1080

aagtacggga aggcgaagaa aagaatagag aagataggga aattagaaga taaaaacata 1140

cttttagaag aaaaaagata aatttaaacc tgaaaagtag gaagcagaag agaaaagaca 1200

agctaggaaa caaaaagcta agggcaaaat gtacaaactt agaagaaaat tggaagatag 1260

aaacaagata gaaaatgaaa atattgtcaa gagtttcaga tagaaaatga aaaacaagct 1320

aagacaagta ttggagaagt atagaagata gaaaaatata aagccaaaaa ttggataaaa 1380

tagcactgaa aaaatgagga aattattggt aaccaattta ttttaaaagc ccatcaattt 1440

aatttctggt ggtgcagaag ttagaaggta aagcttgaga agatgagggt gtttacgtag 1500

accagaacca atttagaaga atacttgaag ctagaagggg aagttggtta aaaatcacat 1560

caaaaagcta ctaaaaggac tggtgtaatt taaaaaaaac taaggcagaa ggcttttgga 1620

agagttagaa gaatttggaa ggccttaaat atagtagctt agtttgaaaa atgtgaagga 1680

ctttcgtaac ggaagtaatt caagatcaag agtaattacc aacttaatgt ttttgcattg 1740

gactttgagt taagattatt ttttaaatcc tgaggactag cattaattga cagctgaccc 1800

aggtgctaca cagaagtgga ttcagtgaat ctaggaagac agcagcagac aggattccag 1860

gaaccagtgt ttgatgaagc taggactgag gagcaagcga gcaagcagca gttcgtggtg 1920

aagataggaa aagagtccag gagccagtgc gatttggtga aggaagctag gaagaaggaa 1980

-continued

ggagcgctaa cgatttggtg gtgaagctag gaaaaaggat tccaggaagg agcgagtgca 2040

atttggtgat gaaggtagca ggcggcttgg cttggcaacc acacggagga ggcgagcagg 2100

cgttgtgcgt agaggatcct agaccagcat gccagtgtgc caaggccaca gggaaagcga 2160

gtggttggta aaaatccgtg aggtcggcaa tatgttgttt ttctggaact tacttatggt 2220

aacctttttat ttattttcta atataatggg ggagtttcgt actgaggtgt aaagggattt 2280

atatggggac gtaggccgat ttccgggtgt tgtaggtttc tctttttcag gcttatactc 2340

atgaatcttg tctgaagctt ttgagggcag actgccaagt cctggagaaa tagtagatgg 2400

caagtttgtg ggtttttttt ttttacacga atttgaggaa aaccaaatga atttgatagc 2460

caaattgaga caatttcagc aaatctgtaa gcagtttgta tgtttagttg gggtaatgaa 2520

gtatttcagt tttgtgaata gatgacctgt ttttacttcc tcaccctgaa ttcgttttgt 2580

aaatgtagag tttggatgtg taactgaggc gggggggagt tttcagtatt tttttttgtg 2640

ggggtggggg caaaatatgt tttcagttct ttttcccttа ggtctgtcta gaatcctaaa 2700

ggcaaatgac tcaaggtgta acagaaaaca agaaaatcca atatcaggat aatcagacca 2760

ccacaggttt acagtttata gaaactagag cagttctcac gttgaggtct gtggaagaga 2820

tgtccattgg agaaatggct ggtagttact ctttttttccc cccacccccт taatcagact 2880

ttaaaagtgc ttaacccctt aaacttgtta ttttttactt gaagcatttt gggatggtct 2940

taacagggaa gagagagggt gggggagaaa atgtttttttt ctaagatttt ccacagatgc 3000

tatagtacta ttgacaaact gggttagaga aggagtgtac cgctgtgctg ttggcacgaa 3060

caccttcagg gactggagct gcttttatcc ttggaagagt attcccagtt gaagctgaaa 3120

agtacagcac agtgcagctt tggttcatat tcagtcatct caggagaact tcagaagagc 3180

ttgagtaggc caaatgttga agttaagttt tccaataatg tgacttctta aaagttttat 3240

taaaggggag gggcaaatat tggcaattag ttggcagtgg cgtgttacgg tgggattggt 3300

ggggtgggtt taggtaattg tttagtttat gattgcagat aaactcatgc cagagaactt 3360

aaagtcttag aatggaaaaa gtaaagaaat atcaacttcc aagttggcaa gtaactccca 3420

atgatttagt ttttttcccc ccagtttgaa ttgggaagct gggggaagtt aaatatgagc 3480

cactgggtgt accagtgcat taatttgggc aaggaaagtg tcataatttg atactgtatc 3540

tgttttcctt caaagtatag agcttttggg gaaggaaagt attgaactgg gggttggtct 3600

ggcctactgg gctgacatta actacaatta tgggaaatgc aaaagttgtt tggatatggt 3660

agtgtgtggt tctcttttgg aatttttttc aggtgattta ataataattt aaaactacta 3720

tagaaactgc agagcaaagg aagtggctta atgatcctga agggatttct tctgatggta 3780

gcttttgtat tatcaaactt ttttcagata acatcttctg agtcataacc agcctggcag 3840

tatgatggcc tagatgcaga gaaaacagct ccttggtgaa ttgataagta aaggcagaaa 3900

agattatatg tcatacctcc attggggaat aagcataacc ctgagattct tactactgat 3960

gagaacatta tctgcatatg ccaaaaaatt ttaagcaaat gaaagctacc aatttaaagt 4020

tacggaatct accattttaa agttaattgc ttgtcaagct ataaccacaa aaataatgaa 4080

ttgatgagaa atacaatgaa gaggcaatgt ccatctcaaa atactgcttt tacaaaagca 4140

gaataaaagc gaaaagaaat gaaaatgtta cactacatta atcctggaat aaaagaagcc 4200

gaaataaatg agagatgagt tgggatcaag tggattgagg aggctgtgct gtgtgccaat 4260

-continued

gtttcgtttg cctcagacag gtatctcttc gttatcagaa gagttgcttc atttcatctg 4320

ggagcagaaa acagcaggca gctgttaaca gataagttta acttgcatct gcagtattgc 4380

atgttaggga taagtgctta tttttaagag ctgtggagtt cttaaatatc aaccatggca 4440

ctttctcctg accccttccc taggggattt caggattgag aaatttttcc atcgagcctt 4500

tttaaaattg taggacttgt tcctgtgggc ttcagtgatg ggatagtaca cttcactcag 4560

aggcatttgc atctttaaat aatttcttaa aagcctctaa agtgatcagt gccttgatgc 4620

caactaagga aatttgttta gcattgaatc tctgaaggct ctatgaaagg aatagcatga 4680

tgtgctgtta gaatcagatg ttactgctaa aatttacatg ttgtgatgta aattgtgtag 4740

aaaaccatta aatcattcaa aataataaac tatttttatt agagaatgta tacttttaga 4800

aagctgtctc cttatttaaa taaaatagtg tttgtctgta gttcagtgtt ggggcaatct 4860

tgggggggat tcttctctaa tctttcagaa actttgtctg cgaacactct ttaatggacc 4920

agatcaggat ttgagcggaa gaacgaatgt aactttaagg caggaaagac aaatttttatt 4980

cttcataaag tgatgagcat ataataattc caggcacatg gcaatagagg ccctctaaat 5040

aaggaataaa taacctctta gacaggtggg agattatgat cagagtaaaa ggtaattaca 5100

cattttattt ccagaaagtc aggggtctat aaattgacag tgattagagt aatacttttt 5160

cacatttcca aagtttgcat gttaacttta aatgcttaca atcttagagt ggtaggcaat 5220

gttttacact attgacctta tatagggaag ggaggggggtg cctgtggggt tttaaagaat 5280

tttcctttgc agaggcattt catccttcat gaagccattc aggattttga attgcatatg 5340

agtgcttggc tcttccttct gttctagtga gtgtatgaga ccttgcagtg agtttatcag 5400

catactcaaa attttttttcc tggaatttgg agggatggga ggaggggggtg gggcttactt 5460

gttgtagctt ttttttttttt tacagacttc acagagaatg cagttgtctt gacttcaggt 5520

ctgtctgttc tgttggcaag taaatgcagt actgttctga tcccgctgct attagaatgc 5580

attgtgaaac gactggagta tgattaaaag ttgtgttccc caatgcttgg agtagtgatt 5640

gttgaaggaa aaaatccagc tgagtgataa aggctgagtg ttgaggaaat ttctgcagtt 5700

ttaagcagtc gtatttgtga ttgaagctga gtacattttg ctggtgtatt tttaggtaaa 5760

atgctttttg ttcatttctg gtggtgggag gggactgaag cctttagtct tttccagatg 5820

caaccttaaa atcagtgaca agaaacattc caaacaagca acagtcttca agaaattaaa 5880

ctggcaagtg gaaatgttta aacagttcag tgatctttag tgcattgttt atgtgtgggt 5940

ttctctctcc cctcccttgg tcttaattct tacatgcagg aacactcagc agacacacgt 6000

atgcgaaggg ccagagaagc cagacccagt aagaaaaaat agcctattta ctttaaataa 6060

accaaacatt ccattttaaa tgtggggatt gggaaccact agttctttca gatggtattc 6120

ttcagactat agaaggagct tccagttgaa ttcaccagtg gacaaaatga ggaaaacagg 6180

tgaacaagct ttttctgtat ttacatacaa agtcagatca gttatgggac aatagtattg 6240

aatagatttc agctttatgc tggagtaact ggcatgtgag caaactgtgt tggcgtgggg 6300

gtggaggggt gaggtgggcg ctaagccttt ttttaagatt tttcaggtac ccctcactaa 6360

aggcaccgaa ggcttaaagt aggacaacca tggagccttc ctgtggcagg agagacaaca 6420

aagcgctatt atcctaaggt caagagaagt gtcagcctca cctgattttt attagtaatg 6480

aggacttgcc tcaactccct ctttctggag tgaagcatcc gaaggaatgc ttgaagtacc 6540

-continued

cctgggcttc tcttaacatt taagcaagct gtttttatag cagctcttaa taataaagcc 6600

caaatctcaa gcggtgcttg aaggggaggg aaaggggaa agcgggcaac cacttttccc 6660

tagcttttcc agaagcctgt taaaagcaag gtctccccac aagcaacttc tctgccacat 6720

cgccaccccg tgccttttga tctagcacag acccttcacc cctcacctcg atgcagccag 6780

tagcttggat ccttgtgggc atgatccata atcggtttca aggtaacgat ggtgtcgagg 6840

tctttggtgg gttgaactat gttagaaaag gccattaatt tgcctgcaaa ttgttaacag 6900

aagggtatta aaaccacagc taagtagctc tattataata cttatccagt gactaaaacc 6960

aacttaaacc agtaagtgga gaaataacat gttcaagaac tgtaatgctg ggtgggaaca 7020

tgtaacttgt agactggaga agataggcat ttgagtggct gagagggctt ttgggtggga 7080

atgcaaaaat tctctgctaa gacttttttca ggtgaacata acagacttgg ccaagctagc 7140

atcttagcgg aagctgatct ccaatgctct tcagtagggt catgaaggtt tttcttttcc 7200

tgagaaaaca acacgtattg ttttctcagg ttttgctttt tggccttttt ctagcttaaa 7260

aaaaaaaaaa gcaaaagatg ctggtggttg gcactcctgg tttccaggac ggggttcaaa 7320

tccctgcggt gtctttgctt tgactactaa tctgtcttca ggactctttc tgtatttctc 7380

cttttctctg caggtgctag ttcttggagt tttggggagg tgggaggtaa cagcacaata 7440

tctttgaact atatacatcc ttgatgtata atttgtcagg agcttgactt gattgtatat 7500

tcatatttac acgagaacct aatataactg ccttgtcttt ttcaggtaat agcctgcagc 7560

tggtgttttg agaagcccta ctgctgaaaa cttaacaatt ttgtgtaata aaaatggaga 7620

agctctaaat tgttgtggtt cttttggaat aaaaaaatct tgattgggaa aaaagatggg 7680

tgttctgtgg gcttgttctg ttaaatctgt ggtctataaa cacagcaccc ataattacag 7740

cataatcttc aagtagggta cggactttgg gggattggtg cgagggtagt gggtgagtgg 7800

cctactaaaa agcccagtaa cccccacagg aaaatagggа acttcttttt aagtagcctc 7860

ctttccacta tttagtaatt ggctgtgagc tgggctgggg gagaaatggg gcggggtgtg 7920

tgtgtcattg gaaagctctc ttttttgttt ttttgagaca gtctcacttt gtcccccagg 7980

ctggagtgta gtggcatgat ctctgcaaac tgcaacctcc acttgtgggg tccaagtggt 8040

tgtcctgctt caccctccct gtagctggga ctacaggtgc acaccaccac gcctggctaa 8100

tttttgtatt 8110

<210> SEQ ID NO 22
<211> LENGTH: 1927
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 22

actttaagtc ttttcctttt ggtcttggag gaacagtgct cagcttttga tctctgtgat 60

tctgatagag aagctgtttc tggaattaca aagtttgcaa acttgacagc aaaatttctg 120

gaaagccttt tctatgaacc tcggacaagt ctaatctgga gctgatcact ctaacatgca 180

cctgtgtggc tgcgactctc ttctggctcc tattaaccct ctttatccga aaaatgaaaa 240

ggtcttcttc tgaaataaag actgactacc tatcaattat aatggaccca gatgaagttc 300

ctttggatga gcagtgtgag cggctccctt atgatgccag caagtgggag tttgcccggg 360

agagacttaa actgggcaaa tcacttggaa gaggggcttt tggaaaagtg gttcaagcat 420

-continued

cagcatttgg cattaagaaa tcacctacgt gccggactgt ggctgtgaaa atgctgaaag 480

aggggccac ggccagcgag tacaaagctc tgatgactga gctaaaaatc ttgacccaca 540

ttggccacca tctgaacgtg gttaacctgc tgggagcctg caccaagcaa ggagggcctc 600

tgatggtgat tgttgaatac tgcaaatatg gaaatctctc caactacctc aagagcaaac 660

gtgacttatt ttttctcaac aaggatgcag cactacacat ggagcctaag aaagaaaaaa 720

tggagccagg cctggaacaa ggcaagaaac caagactaga tagcgtcacc agcagcgaaa 780

gctttgcgag ctccggcttt caggaagata aaagtctgag tgatgttgag gaagaggagg 840

attctgacgg tttctacaag gagcccatca ctatggaaga tctgatttct tacagttttc 900

aagtggccag aggcatggag ttcctgtctt ccagaaagtg cattcatcgg gacctggcag 960

cgagaaacat ttttttatct gagaacaacg tggtgaagat ttgtgatttt ggccttgccc 1020

gggatattta taagaacccc gattatgtga gaaaaggaga tactcgactt cctctgaaat 1080

ggatggctcc tgaatctatc tttgacaaaa tctacagcac caagagcgac gtgtggtctt 1140

acggagtatt gctgtgggaa atcttctcct taggtgggtc tccataccca ggagtacaaa 1200

tggatgagga cttttgcagt cgcctgaggg aaggcatgag gatgagagct cctgagtact 1260

ctactcctga aatctatcag atcatgctgg actgctggca cagagaccca aaagaaaggc 1320

caagatttgc agaacttgtg gaaaaactag gtgatttgct tcaagcaaat gtacaacagg 1380

atggtaaaga ctacatccca atcaatgcca tactgacagg aaatagtggg tttacatact 1440

caactcctgc cttctctgag gacttcttca aggaaagtat ttcagctccg aagtttaatt 1500

caggaagctc tgatgatgtc agatatgtaa atgctttcaa gttcatgagc ctggaaagaa 1560

tcaaaacctt tgaagaactt ttaccgaatg ccacctccat gtttgatgac taccagggcg 1620

acagcagcac tctgttggcc tctcccatgc tgaagcgctt cacctggact gacagcaaac 1680

ccaaggcctc gctcaagatt gacttgagag taaccagtaa aagtaaggag tcggggctgt 1740

ctgatgtcag caggcccagt ttctgccatt ccagctgtgg gcacgtcagc gaaggcaagc 1800

gcaggttcac ctacgaccac gctgagctgg aaaggaaaat cgcgtgctgc tccccgcccc 1860

cagactacaa ctcggtggtc ctgtactcca ccccacccat ctagagtttg acacgaagcc 1920

ttatttc 1927

<210> SEQ ID NO 23
<211> LENGTH: 914
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 23

ctctactgtt agtattgcaa agtatgtatg ggaaacacag agaattggag ctgcgttgaa 60

tgcaaacttg aggtgtttcc cttgaggaat tcttgtcttc aaacgtctgc agagtaatgg 120

accatgttac aactttcctg ttcatctgtg aaccatgaaa atggatggca ctgatgcatt 180

agaccctcag cagcctgcaa ttgcaaatct gcgaggtttc attcggccca taaagcaaac 240

atttgaactt acacagaatg agcacttaaa tacgggtgca ataaatgaag ggaaaaacct 300

cagccgtttc tccattctga agatatagca agcaccggga aatctaagat ttttcatcaa 360

caatatcttc tgccagccca gtttgggggg aaaaacccct tttacatttt tcttcagtaa 420

agtactggaa cttacttttc cctgtctgtg ctaatgagct gattttcagc tgatagaaaa 480

-continued

caaaatgata gagtactttt ttccttggcc aagtattttc tcatttgtat ttaatttcat 540

acattagaca gccagtgaaa ttagacctca aactaggtcc tgatggataa tgaatgttat 600

gtcaccttta acagtgaagt ggttattata ggtcactttc taatttcata ttttcccttt 660

tgctttctgc tgccctcagg gtatatagtg tatctctaac ctgatttttc aaggttattt 720

ttggagcagt ttcttaaaac aggcattccc taacttgctc atttaattaa tgaaaaattg 780

aactgatgcc atggatataa aaacaaatgc aatgtttgat tgtcagtgtt tctgatttgg 840

caaaaaggaa tcatctctat ttttttgcaa acaatatcaa agtgcatatt ttctctcaaa 900

aaaaaaaaaa aaaa 914

<210> SEQ ID NO 24
<211> LENGTH: 3823
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 24

aaagcggggc gcaccgcggg cgccggcaac gagccggtga acgaggcgag gcccgtgcgc 60

ccgcggctgc aagcgcccgc ctggcgggga gaggggccga cggcgtcagc ccgggcggcg 120

gcatccctag gcgcctgggg cgccttcctc cggacctggc cgctcgctgc cccgccctct 180

gcaccccact tctccgaccc tccttcccag tcctgcctcc ccctgccctg gcctctgaga 240

gccgactgag cccagccccg tgcagcagcg gttgcctgtg tcgccgccta gtctccggtc 300

ttggtgctct cccgggggtg ccccaaggag ccagtgcgcg ctgcgggctg ggaaggaggc 360

gccgctcagc tagtcctcct cctcctcctc gtctttctcc tcctcctgct gctgctgccg 420

ccgccgccgc cgtgggtgcc gggtccgcgc gcaccccaac acccccacca gctgggcctc 480

ggggcacatc tccagacttc aatttggctg aaataattca tgccacggac ctgtgcacat 540

gcctggaatt gagagacaca gttaaaagac tccaagttgc tttctgcctt ttgaaaactc 600

ctgaaaacca tccctttgga ctctggaatt ctacacagct caaccaagac tttgcttgaa 660

tgtttacatt ttctgctcgc tgtcctacat atcacaatat agtgttcacg ttttgttaaa 720

actttggggt gtcaggagtt gagcttgctc agcaagccag catggctagg atgagctttg 780

ttatagcagc ttgccaattg gtgctgggcc tactaatgac ttcattaacc gagtcttcca 840

tacagaatag tgagtgtcca caactttgcg tatgtgaaat tcgtccctgg tttaccccac 900

agtcaactta cagagaagcc accactgttg attgcaatga cctccgctta acaaggattc 960

ccagtaacct ctctagtgac acacaagtgc ttctcttaca gagcaataac atcgcaaaga 1020

ctgtggatga gctgcagcag cttttcaact tgactgaact agatttctcc caaaacaact 1080

ttactaacat taaggaggtc gggctggcaa acctaaccca gctcacaacg ctgcatttgg 1140

aggaaaatca gattaccgag atgactgatt actgtctaca agacctcagc aaccttcaag 1200

aactctacat caaccacaac caaattagca ctatttctgc tcatgctttt gcaggcttaa 1260

aaaatctatt aaggctccac ctgaactcca acaaattgaa agttattgat agtcgctggt 1320

ttgattctac acccaacctg gaaattctca tgatcggaga aaaccctgtg attggaattc 1380

tggatatgaa cttcaaaccc ctcgcaaatt tgagaagctt agttttggca ggaatgtatc 1440

tcactgatat tcctggaaat gctttggtgg gtctggatag ccttgagagc ctgtcttttt 1500

atgataacaa actggttaaa gtccctcaac ttgccctgca aaaagttcca aatttgaaat 1560

-continued

tcttagacct caacaaaaac cccattcaca aaatccaaga aggggacttc aaaaatatgc 1620

ttcggttaaa agaactggga atcaacaata tgggcgagct cgtttctgtc gaccgctatg 1680

ccctggataa cttgcctgaa ctcacaaagc tggaagccac caataaccct aaactctctt 1740

acatccaccg cttggctttc cgaagtgtcc ctgctctgga aagcttgatg ctgaacaaca 1800

atgccttgaa tgccatttac caaaagacag tcgaatccct ccccaatctg cgtgagatca 1860

gtatccatag caatcccctc aggtgtgact gtgtgatcca ctggattaac tccaacaaaa 1920

ccaacatccg cttcatggag cccctgtcca tgttctgtgc catgccgccc gaatataaag 1980

ggcaccaggt gaaggaagtt ttaatccagg attcgagtga acagtgcctc ccaatgatat 2040

ctcacgacag cttcccaaat cgtttaaacg tggatatcgg cacgacggtt ttcctagact 2100

gtcgagccat ggctgagcca gaacctgaaa tttactgggt cactcccatt ggaaataaga 2160

taactgtgga aacccttcca gataaataca agctaagtag cgaaggtacc ttggaaatat 2220

ctaacataca aattgaagac tcaggaagat acacatgtgt tgcccagaat gtccaagggg 2280

cagacactcg ggtggcaaca attaaggtta atgggaccct tctggatggt acccaggtgc 2340

taaaaatata cgtcaagcag acagaatccc attccatctt agtgtcctgg aaagttaatt 2400

ccaatgtcat gacgtcaaac ttaaaatggt cgtctgccac catgaagatt gataaccctc 2460

acataacata tactgccagg gtcccagtcg atgtccatga atacaaccta acgcatctgc 2520

agccttccac agattatgaa gtgtgtctca cagtgtccaa tattcatcag cagactcaaa 2580

agtcatgcgt aaatgtcaca accaaaaatg ccgccttcgc agtggacatc tctgatcaag 2640

aaaccagtac agcccttgct gcagtaatgg ggtctatgtt tgccgtcatt agccttgcgt 2700

ccattgctgt gtactttgcc aaaagattta agagaaaaaa ctaccaccac tcattaaaaa 2760

agtatatgca aaaaacctct tcaatcccac taaatgagct gtacccacca ctcattaacc 2820

tctgggaagg tgacagcgag aaagacaaag atggttctgc agacaccaag ccaacccagg 2880

tcgacacatc cagaagctat tacatgtggt aactcagagg atattttgct tctggtagta 2940

aggagcacaa agacgttttt gctttattct gcaaaagtga acaagttgaa gacttttgta 3000

tttttgactt tgctagtttg tggcagagtg gagaggacgg gtggatattt caaatttttt 3060

tagtatagcg tatcgcaagg gtttgacacg gctgccagcg actctaggct tccagtctgt 3120

gtttggtttt tattcttatc attattatga ttgttattat attattattt tattttagtt 3180

gttgtgctaa actcaataat gctgttctaa ctacagtgct caataaaatg attaatgaca 3240

ggatggggtt cccctgtgct tttaccagta gcatgacccc ttctgaagcc atccgtagaa 3300

agtactttgt cctccaaaaa gctaacatac ggttttgaag cagcattgaa acttttgtag 3360

caatctggtc tatagacttt taactcaaga agctaaggct agacttgtta ccttcgttga 3420

atgatgttag ttgactgtac tgtaatgttg tatcaactga attgaatgtt tgcctttaaa 3480

caatgaattt tctttttctt tccttttttt ttttttgtt gtaatagtta aagaggctta 3540

gaacaagcta acaggcaata gaaatatgta tatcagattt tttaatgtaa caaactacat 3600

gttaattgtt atcttattct ttttatcttt agtagacact tttaaaagaa aagacaagtt 3660

tgttgtgttt aactcaccaa cacgtggtgt ataatgaaga cagaactata ataaattagt 3720

tttgttctga ttttttagaa cacttgcaat aatgtatcat ttatagttct tgctagttgc 3780

agtggtaata tttttcacat ccataaaaac aactaccaaa ata 3823

-continued

<210> SEQ ID NO 25
<211> LENGTH: 993
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 25

tcttctccat ttcaaagaca ctgcatatcc ttagaaaccc ttcattacag agggttagtg 60

ttaaaaacct cagtggtagg gtatatatta tataagttaa aaaataaaaa agaagaagcc 120

ctctaaggct caagcagaaa tattctacag aggggaattt tttaggcaga aagaaacagc 180

taggtaagca acccataact ttctccccca ttgctttttc tcacttcagg aggcccagta 240

tggtggtgag aagaaccaga aggcactcaa aatttaggcc agaaggatac gggtatgagg 300

agaaagagat aatccttccc acctgcatgc ttcctcttag gactttaggt ctgcctacat 360

gttgaagaag tttgagtttt aggctttggg agcatttcct taaaaaccat tgtcatgggg 420

gaggaaaaaa aaagggtagt gttaattctg tcttcacaca ctcagagatc atttggttgg 480

gctacatttt atgagctata ttccataaag atgctgcaaa gttcaggtat ctggtttagg 540

aaagaaaaaa aaatcccttc tttaacccaa gaaatgtaat gcaaacaact tgggcctgtg 600

tcacttagtc ttctaagaag acttcaaggc tgttataaag acccagtggt gggagaaacc 660

tgatccggta tagcggggcc agaagaagct gcgggaccat gccgcgggga caaggcaagg 720

aaatttgaac aataacagct gctttcagaa ataaaccgcg actttgtgca gcgtgtatta 780

cctgacgagc tgaaatgcat tgttaatgag actcgcccga tcattactgc tgactgctgt 840

gtgtgttcct tttaaaaggc cagtcaaaga gtcccatcca tcatcctcgt aatgcacaat 900

gtaatagcca ttcatgccca cattaaattt gatccattcc acctcttctg ggaggatgag 960

cacatctaga gtaaataaaa taaatcaaag ttc 993

<210> SEQ ID NO 26
<211> LENGTH: 5584
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 26

acggatccgc gttcagaaag gcgtgcactt cctacgcctg atcccccgca tcgcaacctc 60

gcagcttccc cggcgtgcag cgctcattta ccaattccct tcctgggagt tgcggcttcc 120

ctcgctcggc cccactcccg tttacccttt ccccagctcc cgccttagcc aggggcttcc 180

ccgcctgccg ctagggctcg ggccgaagcg ccgctcagcg ccagcctgcc gctccccggg 240

ctccactttc actttcggtc ctgggggagc taggccggcg gcagtggtgg tggcggcggc 300

gcaagggtga gggcggcccc agaaccccag gtaggtagag caagaagatg gtgtttctgc 360

ccctcaaatg gtcccttgca accatgtcat ttctactttc ctcactgttg gctctcttaa 420

ctgtgtccac tccttcatgg tgtcagagca ctgaagcatc tccaaaacgt agtgatggga 480

caccatttcc ttggaataaa atacgacttc ctgagtacgt catcccagtt cattatgatc 540

tcttgatcca tgcaaacctt accacgctga ccttctgggg aaccacgaaa gtagaaatca 600

cagccagtca gcccaccagc accatcatcc tgcatagtca ccacctgcag atatctaggg 660

ccaccctcag gaagggagct ggagagaggc tatcggaaga acccctgcag gtcctggaac 720

acccccgtca ggagcaaatt gcactgctgg ctcccgagcc cctccttgtc gggctcccgt 780

-continued

acacagttgt cattcactat gctggcaatc tttcggagac tttccacgga ttttacaaaa 840

gcacctacag aaccaaggaa ggggaactga ggatactagc atcaacacaa tttgaaccca 900

ctgcagctag aatggccttt ccctgctttg atgaacctgc cttcaaagca agtttctcaa 960

tcaaaattag aagagagcca aggcacctag ccatctccaa tatgccattg gtgaaatctg 1020

tgactgttgc tgaaggactc atagaagacc attttgatgt cactgtgaag atgagcacct 1080

atctggtggc cttcatcatt tcagattttg agtctgtcag caagataacc aagagtggag 1140

tcaaggtttc tgtttatgct gtgccagaca agataaatca agcagattat gcactggatg 1200

ctgcggtgac tcttctagaa ttttatgagg attatttcag cataccgtat cccctaccca 1260

aacaagatct tgctgctatt cccgactttc agtctggtgc tatggaaaac tggggactga 1320

caacatatag agaatctgct ctgttgtttg atgcagaaaa gtcttctgca tcaagtaagc 1380

ttggcatcac aatgactgtg gcccatgaac tggctcacca gtggtttggg aacctggtca 1440

ctatggaatg gtggaatgat ctttggctaa atgaaggatt tgccaaattt atggagtttg 1500

tgtctgtcag tgtgacccat cctgaactga aagttggaga ttatttcttt ggcaaatgtt 1560

ttgacgcaat ggaggtagat gctttaaatt cctcacaccc tgtgtctaca cctgtggaaa 1620

atcctgctca gatccgggag atgtttgatg atgtttctta tgataaggga gcttgtattc 1680

tgaatatgct aagggagtat cttagtgctg acgcatttaa aagtggtatt gtacagtatc 1740

tccagaagca tagctataaa aatacaaaaa acgaggacct gtgggatagt atggcaagta 1800

tttgccctac agatggtgta aaagggatgg atggcttttg ctctagaagt caacattcat 1860

cttcatcctc acattggcat caggaagggg tggatgtgaa aaccatgatg aacacttgga 1920

cactgcagaa gggttttccc ctaataacca tcacagtgag ggggaggaat gtacacatga 1980

agcaagagca ctacatgaag ggctctgacg gcgccccgga cactgggtac ctgtggcatg 2040

ttccattgac attcatcacc agcaaatccg acatggtcca tcgatttttg ctaaaaacaa 2100

aaacagatgt gctcatcctc ccagaagagg tggaatggat caaatttaat gtgggcatga 2160

atggctatta cattgtgcat tacgaggatg atggatggga ctctttgact ggccttttaa 2220

aaggaacaca cacagcagtc agcagtaatg atcgggcgag tctcattaac aatgcatttc 2280

agctcgtcag cattgggaag ctgtccattg aaaaggcctt ggatttatcc ctgtacttga 2340

aacatgaaac tgaaattatg cccgtgtttc aaggtttgaa tgagctgatt cctatgtata 2400

agttaatgga gaaaagagat atgaatgaag tggaaactca attcaaggcc ttcctcatca 2460

ggctgctaag ggacctcatt gataagcaga catggacaga cgagggctca gtctcagagc 2520

gaatgctgcg gagtcaacta ctactcctcg cctgtgtgca caactatcag ccgtgcgtac 2580

agagggcaga aggctatttc agaaagtgga aggaatccaa tggaaacttg agcctgcctg 2640

tcgacgtgac cttggcagtg tttgctgtgg gggcccagag cacagaaggc tgggattttc 2700

tttatagtaa atatcagttt tctttgtcca gtactgagaa aagccaaatt gaatttgccc 2760

tctgcagaac ccaaaataag gaaaagcttc aatggctact agatgaaagc tttaagggag 2820

ataaaataaa aactcaggag tttccacaaa ttcttacact cattggcagg aacccagtag 2880

gatacccact ggcctggcaa tttctgagga aaaactggaa caaacttgta caaaagtttg 2940

aacttggctc atcttccata gcccacatgg taatgggtac aacaaatcaa ttctccacaa 3000

gaacacggct tgaagaggta aaaggattct tcagctcttt gaaagaaaat ggttctcagc 3060

-continued

tccgttgtgt ccaacagaca attgaaacca ttgaagaaaa catcggttgg atggataaga 3120

attttgataa aatcagagtg tggctgcaaa gtgaaaagct tgaacatgat cctgaagctg 3180

acgcaacagg atgaaaatcc atcagaatct cagactacag cactaaatat gctttgatgc 3240

tacatcaaac ggaatggaag catagctgac ttcgctaaag ttacttcatc tccatctagc 3300

aaatgaggca ctgttctcaa ccaaaggaga tggggatctg gtttagggca atccctttat 3360

aatttgatgt gctgtggtct ccttggtaat gtataatttg gtattgcaca ggtgattagt 3420

caaggaagtc tggaaaagct ttggtcccac agccttgcct cacagcatgt aaataattaa 3480

aacaatattg atgctgaggt tcttctactg ctagtatgaa agtgacaaat ttttactggt 3540

gtgaattggg aagaaaacaa tgctattcca tgacgtttgt aaaatgtttg taaaagctca 3600

aacatgacga ttccataaaa taaacttgag gttaaataat gggtagtaaa ttatagaatg 3660

tataagaaaa aatataaagg agaaaatcaa ttatcaggaa agctaaagaa cttttcaaat 3720

ctagtaattt gaatatagac acaatgcact ttattgcact ttcaattctt ataaagcaac 3780

aataatatta aggtccttga ctatgtgtac aatgttttca catatatagt ttcatttaat 3840

catttcaaag ttaatctctg ccatctcgct aaatcatcag tctcggctct tctgaaatag 3900

aaggtgcctg atcttcctaa taattctgcc tattttcatt tgctttaaac aggcgcccta 3960

ttttctttct agttgtggct gcgcaaaaac atttatctcc caaataagat gtgctgctta 4020

ccgaggtatc acggggtggg gctccagctt gggtcgttga agctggggtt tgggaaacca 4080

cttcagagat ggcagcagca agtttagcat cttcaaattt cttttattga aaaaaatttt 4140

attagtaaca tgttgtatat aaaattatga gcacaatgcc atcacttaac tataactctt 4200

aaagatagct taatgactgt ttattctctt gaccaaatag actcataata acatataatt 4260

ttaaaagaaa tttaaattct ttcttctcta ttgtattatt ttatacaatt tgctatttct 4320

atttccttct catattgatt attctaaata ctatgcaata atataactta gagttccacg 4380

gtttgtttac acatttcctg ttgtacattt aggttattca aagttttcag ctcttttaaa 4440

attgctctga ataagttcta gtgagtgagt tatggtgctg gctatatttt gctaaactgc 4500

cctctcaaat gttgctagga attcatactg cgaaaagcaa tgaataagca tgcctgtttt 4560

cccatggcct tgcttgccag aatttgactt ttattatgat aatcagtgta aaatgatata 4620

ctactattgc ttgtatattg tggtatacgg tgtcaggttt cagggttttt tttcaacgtt 4680

aaatattcta gaaactttct gaaataattt ctgtttaaaa atattgaata tttgcttcat 4740

ttcaaatact ccctttttgac aaaaaaactt aggtataact gttgatgaaa aaccagaaaa 4800

aagtccagaa ctctttggtg actccaacta tggatagctt attttgaaaa aggagaattg 4860

caaattttac caaaagatgg agaaaagcac attaaaaaga taccaacatt cagaaattca 4920

tttcagcatg ttattattgg aaattattta aactaattta gataactata agatacttat 4980

tgtccattta taccctgtaa agccgtttta gaatgtaata ttttaggtaa tccaaaatgt 5040

actaaattaa attcattttt agttatgaga aatctttgct tatatgacaa atgaaaagaa 5100

taacaagttg tcaaatgaaa agaatgacat tgaaacattt gtattgtctc ttcttaaact 5160

atcttattga cttattattt aagcctttta atactaagta tgaaacaacc tatggtctgg 5220

aaatttgtat cgcaaagcta tatgtgcata tgttatttaa ttcatctaat gctacacaaa 5280

agcataaaat aatgattttt cactctcttt aaaaatacta aatcatttat gtccatttct 5340

-continued

caatttttttc attgatctat gctttgagtt tgctttctca acattattgt attttccact 5400

tattattact gtataacata tgctagtgtt tagttggatt aatcttacct aaaagtactg 5460

aaaaatgctt tttagtactt tttcatattt tatacattta ttttccgaat gtatcattga 5520

ataattttat tgagttataa aagtatctta ttgctattta ataaaaaatt aacacataaa 5580

atga 5584

<210> SEQ ID NO 27
<211> LENGTH: 2341
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 27

ggctgcgtgg ataaaaccca aagggaatcc agaatcaagc tcggcctcgc aggagtttat 60

aacttttgta cttttccctc ccttgtaacg aaaagctaat cgcaaacgtt tctaggaaca 120

catttttctc ctgatagacg ttcccaaggt ggagggcgcg agaagactgg cccgtggggt 180

gaggagcccc ggcccgacct cagcggcgcc acaaagccat gaagagggga aataattttc 240

caaggggcta agagttggac tttctgaagc catcgaaacc tgctgctagg ctctcccgct 300

ggggttttga gagaggtagg tgaggaggga tcaaacatct ccacccgctc ttgaaagaac 360

cgccttgagc tgcgcccaac ttctgcaatc ccggttttcc ccaagttcaa ggggaagggc 420

cctggcgccc aatgcccagc ctccccgaca acatcctcgc cagggtgcgc ctgccccgcg 480

ggtccgctgt ccccggcggg cggccctggg ttcgaggacc cggcttcccg gaggcgcgac 540

tctccagcaa acaagcggag ctccagaaaa aacaccgacg cgaagggggt ggacagaggg 600

tgggaagccg cggggttgtc gcccacccgg aggcggaatg taacgcgctg tgggaaacgt 660

ggcaggagcc ctgaggcggc gtctttttcc tcaggcgccc gccgagggtc ccgcggaaag 720

cctcgcgggc cgccctcgcg gcccgcagcc ccgctctggc gccgagccgg agcccatgca 780

acctggttcc atcccctgcc cctcccctgt ccccggtgcg ccgcccgcca gcgagccttc 840

gacgtggccg caggggcgac gggaccaccc tcccccgata cccacagccc cgccatgtct 900

gcctttcccc ggcccggtct cctcaccgcc gctgccgccg ccgctcttcg gccggagatt 960

cggcggccca gaccgtgtcc tggccgggaa ctgagggctc cgcctcaacg ggcccgcgct 1020

gggcaacagg gagcgcagcg agcctcgtcc ggcgcgtgcg gctcccgcgt cgtcgggcgg 1080

ccaagcgcgc gttgagagga ctggcgggcg gacgagcgcg cacacgagtg agcgcagccc 1140

caaagcggcg cgcagggggc tcgcggcccg gaagagggga ggggcgatga cccgggaaag 1200

ggttggcgcg cgcgggatcg gctcgcgcgc ctgaggggcg gtgccggggg cggggcttcg 1260

ccgcgaggcc acccccgagc cccgggccta gccgcacggg aggcgacaca ccctcgccct 1320

accctgagcc tggcgcagac cctcgaccgc gacggcgcgg tccgcagaac cgcgggcttc 1380

cctccccggc aaaaagcggc gctgaaggcc tgtgggaccc ccgcccggcc gccctacgtg 1440

cgtcagcgcc ttcgggggtc gcgcgggctt ggtcctttcg gctgtctcgg ccttttttgtt 1500

tcccccctcag gtctctccac gctgcacttg acaccctcag ggcggaatgg cgagttctag 1560

acccagctct ctagacccgg ggcttcatgg gggacacgga ctcgacggga aggggaactt 1620

ggcatttatg gacagatcct catccttctt gctgtagtca attacacgca cgttaaccgt 1680

gcagccgccc tgctgtattt taggcggttg ttggcactcc tagttgggcc cttccctggc 1740

-continued

ccctcacagc agggcctgcc tcctgtggac gcttgtgtgc tgccctggca ccggccactg 1800

tgttttgcat aaaggagagg ctccaaagat gttggccaaa ggaatgaagc cttgagagtc 1860

caggctttct atttctgaga cactctactc aaggacttcc cagatgcaaa gcttcatctt 1920

tgagcaaata caaatatgga gagggaacat taacttttctg aagaaaagaa ggaacatctt 1980

ttcaaccttt tatattgagt taacaccatg gtcttagttt tgtgaaagca ctttaataca 2040

tccaactagc gggaggaata caagatattc ctacctttttt attattatta ttaaaagagg 2100

gctaaaccat gttgtaaact acttaagaac tgaatgttca agttgtacag tatattggcc 2160

agattcctta ctgcccacta ggcttgggag catgttttca gctcagtttt atgaatgttt 2220

taaatttgta ttactgatct ttgttatagc caattagcaa ttcaaaatgg taatttttttt 2280

agcagtaata aatataccat cccccaaaca gtaaaaaaaa aaaaaaaaaa aaaaaaaaaa 2340

a 2341

<210> SEQ ID NO 28
<211> LENGTH: 1596
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 28

ccctcttccg ggccgcgagc cccctgcgcg ccgctttggg gctgcgctca ctcgtgtgcg 60

cgctcgtccg cccgccagtc ctctcaacgc gcgcttggcc gcccgacgac gcgggagccg 120

cacgcgccgg acgaggctcg ctgcgctccc tgttgcccag cgcgggcccg ttgaggcgga 180

gccctcagtt cccggccagg acacggtctg ggccgccgaa tctccggccg aagagcggcg 240

gcggcagcgg cgggaaaaaa atgaagaatg aaattgctgc cgttgtcttc tttttcacaa 300

ggctagttcg aaaacatgat aagttgaaaa aagaggcagt tgagaggttt gctgagaaat 360

tgaccctaat acttcaagaa aaatataaaa atcactggta tccagaaaaa ccatcgaaag 420

gacaggccta cagatgtatt cgtgtcaata aatttcagag agttgatcct gatgtcctga 480

aagcctgtga aaacagctgc atcttgtata gtgacctggg cttgccaaag gagctcactc 540

tctgggtgga cccatgtgag gtgtgctgtc gtagagatgg ggtttcacca tgttggccag 600

actgctctca aactcctgac ctcgtgatcc gcccgccttg gcctcccaaa gcgctggatt 660

acaggcgtga gccactgcgc ccggcctcct ccttttttgat tatgtatgga gagaaaaaca 720

atgcattcat tgttgccagc tttgaaaata aagatgagaa caaggatgag atctccagga 780

aagttaccag ggcccttgat aaggttacct ctgattatca ttcaggatcc tcttcttcag 840

atgaagaaac aagtaaggaa atggaagtga aacccagttc ggtgactgca gccgcaagtc 900

ctgtgtacca gatttcagaa cttatatttc cacctcttcc aatgtggcac cctttgccca 960

gaaaaaagcc aggaatgtat cgagggaatg gccatcagaa tcactatcct cctcctgttc 1020

catttggtta tccaaatcag ggaagaaaaa ataaaccata tcgcccaatt ccagtgacat 1080

gggtacctcc tcctggaatg cattgtgacc ggaatcactg gattaatcct cacatgttag 1140

cacctcacta acttcgtttt tgattgtgtt ggtgtcatgt tgagaaaaag gtagaataaa 1200

ccttactaca cattaaaagt taaaagttct tactaatagt agtgaagtta gatgggccaa 1260

accatcaaac ttatttttat agaagttatt gagaataatc tttcttaaaa aatatatgca 1320

ctttagatat tgatatagtt tgagaaattt tattaaagtt agtcaagtgc ctaagttttt 1380

-continued

aatattggac ttgagtattt atatattgtg catcaactct gttggatacg agaacactgt 1440

agaagtggac gatttgttct agcacctttg agaatttact ttatggagcg tatgtaagtt 1500

atttatatac aaggaaatct attttatgtc gttgtttaag agaattgtgt gaaatcatgt 1560

agttgcaaat aaaaaatagt ttgaggcatg acaaaa 1596

<210> SEQ ID NO 29
<211> LENGTH: 1810
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 29

tctcccggag gcgcatgatg tcctcggcca ggttgtcgcg ctccacctcg acgcgggctt 60

tgtcgttggt tagctggtcc acctgccggc gcagctcccg catctcctcc tcgtagaggt 120

cccccaggcg cgacttgcct tggcccttga gctgctcgag ctcggccagc aggatcttat 180

tctgctgctc caggaagcgc accttgtcga tgtagttggc gaagcggtca ttcagctcct 240

gcagctccac cttctcgttg gtgcgggtgt tcttgaactc ggtgttgatg gcgtcggcca 300

gcgagaagtc caccgagtcc tgcaggagcc gcaccccggg cacgctgctc cgcaggcgca 360

cggcagagga gcgcgtggca tacacgccgc ccggggacga ggcgtagagg ctgcggctgg 420

tgctggggcg cagcgcgctg cccaggctgt aggtgcgggt ggacgtagtc acgtagctcc 480

ggctggagct cggccggctc gcggtgcccg ggccgccgaa catcctgcgg taggaggacg 540

aggacacgga cctggtggac atggctgcgg agggtggcga tggcctgggc ggcggcggtg 600

gcgcggactg gctcccggag aagaggcgaa cgagggcgcg acagcaaagc tcccttgga 660

tgacatagat ttattactta gtagtatatt atgtattggc tgtcccacat tttgaaattt 720

aatggactcg tggtgatcag aataaaagaa gccttcgatg tgaggcccag gcattgagta 780

ccatgtgtgc gattcacaag ccttggcatc tgaacaattt tttggaagga tccatcagga 840

caacacgtcg ggggtgtact agtgaagtga ttttccaaat gtgctactca gacctgtagc 900

atcagcatca ccggatatac aagacttacc aggtgattct gagaaccact gtgctagtga 960

attgttcctg tctgtggcca cgtagaataa gaaaaccata gggttggaaa atggggaaac 1020

tggtgagtgt tcgcttatga ggaaatgaag aatagataga aaagaattag ttaacttttg 1080

gaattcaaaa gagaagcagt ttgtaaaagc caggaatttg atttgaagga ctaattgctt 1140

gcagaatctt tgctttctca gagaggggca atccagatca ctaggttacc gtgaaatatg 1200

ttggtatggc tctaaaattc tagaataatc tctctagtga gaaaaggcat acctttccat 1260

attaggacaa aacttcaaat caggtttgta aaatcctata agattattgt atcccattgc 1320

catggccaac ttgtttgtct ctggagatcc cagtttctac atctgaaaac catatgcatt 1380

cctgctcacc aggaattatc tcgctgaatg aagcaagaga ttcaggatgt gtaagagaat 1440

ttatgaagca tttggattac caagaaatgt gaagtataaa gatcaagaat gattattaca 1500

aacactgatg gtaaagaggt gtttcgaagt cgatgcaaaa aatgagttgc ttatttcagt 1560

ctctctttga tatatctgcc tttttagtgc tgactctatc atgtattcac tatttgattt 1620

tcagtgaatc acatttttta aagcttttaa tctgtttcct caaaaaatat atttttaaa 1680

aatatttact caggattgtt gtgagaataa aattgattcg attgatattt caaaaaagaa 1740

atatattttt aaaaaataaa gcaataccat ttttggaaaa aaaaaaaaaa aaaaaaaaaa 1800

-continued

aaaaaaaaaa 1810

<210> SEQ ID NO 30
<211> LENGTH: 2151
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 30

gcctctccaa aggctgcaga agtttcttgc taacaaaaag tccgcacatt cgagcaaaga 60

caggctttag cgagttatta aaaacttagg ggcgctcttg tcccccacag ggcccgaccg 120

cacacagcaa ggcgatggcc cagctgtaag ttggtagcac tgagaactag cagcgcgcgc 180

ggagcccgct gagacttgaa tcaatctggt ctaacggttt cccctaaacc gctaggagcc 240

ctcaatcggc gggacagcag ggcgcgtcct ctgccactct cgctccgagg tccccgcgcc 300

agagacgcag ccgcgctccc accacccaca cccaccgcgc cctcgttcgc ctcttctccg 360

ggagccagtc cgcgccaccg ccgccgccca ggccatcgcc accctccgca gccatgtcca 420

ccaggtccgt gtcctcgtcc tcctaccgca ggatgttcgg cggcccgggc accgcgagcc 480

ggccgagctc cagccggagc tacgtgacta cgtccacccg cacctacagc ctgggcagcg 540

cgctgcgccc cagcaccagc cgcagcctct acgcctcgtc cccgggcggc gtgtatgcca 600

cgcgctcctc tgccgtgcgc ctgcggagca gcgtgcccgg ggtgcggctc ctgcaggact 660

cggtggactt ctcgctggcc gacgccatca acaccgagtt caagaacacc cgcaccaacg 720

agaaggtgga gctgcaggag ctgaatgacc gcttcgccaa ctacatcgac aaggtgcgct 780

tcctggagca gcagaataag atcctgctgg ccgagctcga gcagctcaag ggccaaggca 840

agtcgcgcct gggggacctc tacgaggagg agatgcggga gctgcgccgg caggtggacc 900

agctaaccaa cgacaaagcc cgcgtcgagg tggagcgcga caacctggcc gaggacatca 960

tgcgcctccg ggagaaattg caggaggaga tgcttcagag agaggaagcc gaaaacaccc 1020

tgcaatcttt cagacaggat gttgacaatg cgtctctggc acgtcttgac cttgaacgca 1080

aagtggaatc tttgcaagaa gagattgcct ttttgaagaa actccacgaa gaggaaatcc 1140

aggagctgca ggctcagatt caggaacagc atgtccaaat cgatgtggat gtttccaagc 1200

ctgacctcac ggctgccctg cgtgacgtac gtcagcaata tgaaagtgtg gctgccaaga 1260

acctgcagga ggcagaagaa tggtacaaat ccaagtttgc tgacctctct gaggctgcca 1320

accggaacaa tgacgccctg cgccaggcaa agcaggagtc cactgagtac cggagacagg 1380

tgcagtccct cacctgtgaa gtggatgccc ttaaaggaac caatgagtcc ctggaacgcc 1440

agatgcgtga aatggaagag aactttgccg ttgaagctgc taactaccaa gacactattg 1500

gccgcctgca ggatgagatt cagaatatga aggaggaaat ggctcgtcac cttcgtgaat 1560

accaagacct gctcaatgtt aagatggccc ttgacattga gattgccacc tacaggaagc 1620

tgctggaagg cgaggagagc aggatttctc tgcctcttcc aaacttttcc tccctgaacc 1680

tgagggaaac taatctggat tcactccctc tggttgatac ccactcaaaa aggacacttc 1740

tgattaagac ggttgaaact agagatggac aggttatcaa cgaaacttct cagcatcacg 1800

atgaccttga ataaaaattg cacacactca gtgcagcaat atattaccag caagaataaa 1860

aaagaaatcc atatcttaaa gaaacagctt tcaagtgcct ttctgcagtt tttcaggagc 1920

gcaagataga tttggaatag gaataagctc tagttcttaa caaccgacac tcctacaaga 1980

-continued

tttagaaaaa agtttacaac ataatctagt ttacagaaaa atcttgtgct agaatacttt 2040

ttaaaaggta ttttgaatac cattaaaact gctttttttt ttccagcaag tatccaacca 2100

acttggttct gcttcaataa atctttggaa aaactcaaaa aaaaaaaaaa a 2151

<210> SEQ ID NO 31
<211> LENGTH: 590
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 31

caaaatagtc aggaaagtta tgaagtaaca tcatataata ctttatcccc aacaatatta 60

aatgtgatta caattttaca gaatattcaa atcaaaatgt tatataagac aaattattat 120

tcttgctttt acatctcctt agaaggagaa acattcaatt gttttaaaaa tgttctattt 180

atttttcaga aatgaggtgt tgccttcttg cccaggctag tatgcagtgg cacaatcata 240

gctcactgca gccttgaatt cttgggctca agtgatcttc ctgccccaag cctcccaagt 300

agctgggacc acaggcacgt gccaccatgc ctggctgatt ttttttttt tttaatgttt 360

tgtagagaca ggatctcagt atgttgcgca ggctggtctt taattcctgg gctcaagtga 420

tcctcccata gatttgtctg tgttttctct aaataggtta taaatacctt gagggaggaa 480

tcacaaataa ctaaaagaaa cttttgtaga gactatcaga gtattatttc cttgtggaca 540

cagtaaatgt atgatgatga gaaatcataa taaactacaa cattttctcc 590

<210> SEQ ID NO 32
<211> LENGTH: 3050
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 32

tttttttgct tctgccccag atctttcctg gacagtgcgt ctcagcagtt cagatccggg 60

ggcccccagc tgacagaggg cgtgggggggt taaggcatta acccctccca gcctcttcct 120

gaagaaacca cccagccttg gcgcggcgct gggtgacttc gcgtagcagg cagggaactg 180

gccgcggcga gcgggactgg ccattggagt gctccgctgc ggagggaggg gaccccgact 240

cgagtaagtt tgcgagagca ctacgcagtc agtcgggggc agcagcaaga tgcgaagcga 300

gccgtacaga tcccgggctc tccgaacgca acttcgccct gcttgagcga ggctgcggtt 360

tccgaggccc tctccagcca aggaaaagct acacaaaaag cctggatcac tcatcgaacc 420

acccctgaag ccagtgaagg ctctctcgcc tcgccctcta gcgttcgtct ggagtagcgc 480

caccccggct tcctggggac acagggttgg caccatgggg cccaccagcg tcccgctggt 540

caaggcccac cgcagctcgg tctctgacta cgtcaactat gatatcatcg tccggcatta 600

caactacacg ggaaagctga atatcagcgc ggacaaggag aacagcatta aactgacctc 660

ggtggtgttc attctcatct gctgctttat catcctggag aacatctttg tcttgctgac 720

catttggaaa accaagaaat tccaccgacc catgtactat tttattggca atctggccct 780

ctcagacctg ttggcaggag tagcctacac agctaacctg ctcttgtctg gggccaccac 840

ctacaagctc actcccgccc agtggtttct gcgggaaggg agtatgtttg tggccctgtc 900

agcctccgtg ttcagtctcc tcgccatcgc cattgagcgc tatatcacaa tgctgaaaat 960

gaaactccac aacgggagca ataacttccg cctcttcctg ctaatcagcg cctgctgggt 1020

-continued

catctccctc atcctgggtg gcctgcctat catgggctgg aactgcatca gtgcgctgtc 1080

cagctgctcc accgtgctgc cgctctacca caagcactat atcctcttct gcaccacggt 1140

cttcactctg cttctgctct ccatcgtcat tctgtactgc agaatctact ccttggtcag 1200

gactcggagc cgccgcctga cgttccgcaa gaacatttcc aaggccagcc gcagctctga 1260

gaagtcgctg gcgctgctca agaccgtaat tatcgtcctg agcgtcttca tcgcctgctg 1320

ggcaccgctc ttcatcctgc tcctgctgga tgtgggctgc aaggtgaaga cctgtgacat 1380

cctcttcaga gcggagtact tcctggtgtt agctgtgctc aactccggca ccaaccccat 1440

catttacact ctgaccaaca aggagatgcg tcgggccttc atccggatca tgtcctgctg 1500

caagtgcccg agcggagact ctgctggcaa attcaagcga cccatcatcg ccggcatgga 1560

attcagccgc agcaaatcgg acaattcctc ccacccccag aaagacgaag gggacaaccc 1620

agagaccatt atgtcttctg gaaacgtcaa ctcttcttcc tagaactgga agctgtccac 1680

ccaccggaag cgctctttac ttggtcgctg gccaccccag tgtttggaaa aaaatctctg 1740

ggcttcgact gctgccaggg aggagctgct gcaagccaga gggaggaagg gggagaatac 1800

gaacagcctg gtggtgtcgg gtgttggtgg gtagagttag ttcctgtgaa caatgcactg 1860

ggaagggtgg agatcaggtc ccggcctgga atatattttc tacccccctg gagctttgat 1920

tttgcactga gccaaaggtc tagcattgtc aagctcctaa agggttcatt tggcccctcc 1980

tcaaagacta atgtccccat gtgaaagcgt ctctttgtct ggagctttga ggagatgttt 2040

tccttcactt tagtttcaaa cccaagtgag tgtgtgcact tctgcttctt tagggatgcc 2100

ctgtacatcc cacaccccac cctcccttcc cttcataccc ctcctcaacg ttcttttact 2160

ttatacttta actacctgag agttatcaga gctggggttg tggaatgatc gatcatctat 2220

agcaaatagg ctatgttgag tacgtaggct gtgggaagat gaagatggtt tggaggtgta 2280

aaacaatgtc cttcgctgag gccaaagttt ccatgtaagc gggatccgtt ttttggaatt 2340

tggttgaagt cactttgatt tctttaaaaa acatcttttc aatgaaatgt gttaccattt 2400

catatccatt gaagccgaaa tctgcataag gaagcccact ttatctaaat gatattagcc 2460

aggatccttg gtgtcctagg agaaacagac aagcaaaaca aagtgaaaac cgaatggatt 2520

aacttttgca aaccaaggga gatttcttag caaatgagtc taacaaatat gacatctgtc 2580

tttggcactt ttgttgatgt ttatttcaga atgttgtgtg attcatttca agcaacaaca 2640

tggttgtatt ttgttgtgtt aaaagtactt ttcttgattt ttgaatgtat ttgtttcagc 2700

agaagtcatt ttattggatt tttctaaccc gtgttaacac cattgaatgt gtatttctta 2760

agaaaatacc accctcttgt gcccttaaaa gcattacttt aactggtagg gaacgccaga 2820

aacttttcag tccagctatt cattagatag taattgaaga tatgtataaa tattacaaag 2880

aataaaaata tattactgtc tctttagtat ggttttcagt gcaattaaac cgagagatgt 2940

cttgtttttt taaaaagaat agtatttaat aggtttctga cttttgtgga tcattttgca 3000

catagcttta tcaactttta aacattaata aactgatttt tttaaagatc 3050

<210> SEQ ID NO 33
<211> LENGTH: 3186
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 33

-continued

ctagcatggc agcttccccc tcctcgatga agcccagcac ctgggtctgg aagccctgca 60

gggcggccgc agcatctgca aacagccggc tcaccctctc ccgctctgct acggctgcac 120

tctgcacagg acgacagtag agggggcaat gagggcaaag aaccgtcccg gactgtaccc 180

ccctccttcc ctgaccccgg cttcccaccg gccctcttga ggggacagag gggccagact 240

gagcctgtcc tggatctggg ccaggtcagg ggacacttgt gctcagtagg tatgggcatg 300

gtttgtttgg agtggattgt tccatgctga gggaggggtg tttgaccttg atgagggcca 360

ctgtgcgcct ggactgtgca atgccagcac ccagctcgtc catgcggtcc tccacggcgc 420

tcaggacttt ggactgctca gcctgtggac aacaccctcc atgagtgtga ggtctggcag 480

aggccacagc cctacgctcg ggtcccaggc cttgccaggg gagagaaggg ggtgactggg 540

gagtggtatc tgtgcatcag gtgggactgg agtggccacg atgtcccata atatgccaga 600

ggtgcctgca gcccaggctg gtgcctttcc tgaacgcctg gggggcgtgc aggctgactg 660

ggtgtctgcc tcgggtctct gacaggaggc ctagtacagg gccgtccgct ggtctcagtg 720

ttgctcagat gaatttgaga acttcctccg gcggggaaga ggcagagagt aagccagtta 780

ggttgggagg tgggcggcgg aggggtgact gcacaaggag gcattaagct gtcctgagga 840

tcactgttcc caaagagcct gcatccacag tccaggcctg gaaaaaaggg gattccgggc 900

aaatccccta ggtcttatga gtcactgctc caggcaggag aaacggaaaa ccccttggcc 960

agaggcgcca agaccgggct ggaggccaaa ggaaactggg ggcgggaggt agaggccgcg 1020

gctcacatgc tcccaccact ggcccccgga acagctggac agagaatctg tgtgcccggt 1080

cgccctctgc cgccctcctg ccgcctgcca cacccatcga aagctgctgc ctcagcccag 1140

ggcaaatctg agctgagacc ggcgtgaccc tccgggcggg agcagcagcc acgctaccgg 1200

aaccccgcgg gcctaacccg gctccctccc gcctacaccc ccatagaccc cggcccggct 1260

tcaggtcccc tgcccgctct tcgcgtcctt gtcctgttct cacgtccctc cttcctccct 1320

tcccatctcc atcaccgcca gcagatgacc ttgtcaaggc tggaagggac cgcaaagatc 1380

ctcaggttag aggggagat taagacccaa cccgggaaga ctgggggtc caaggtcctc 1440

ggcccgctcg cgcgcggtcg gtcaccagcc ccggctcacc cgagtgtgcc cccggagccc 1500

acctcctgaa gcgcgcgctc ctgctccagc ggcaccagct cgtggccgcg gtgctcctgt 1560

gcggcgcagg cctcgcacag acacacgcgc tccgcgcggc agtagcgctc gagcggccgt 1620

aggtggcgcg ggcacaggct ctcctctagc cggcgcagcg gcggcaccag gcggtgtccg 1680

cggagggcgg ggctgcgctc gtgcgggccc aggtgcgcgg ggcaaaagga ggcgaggcag 1740

gagaggcagg acagcgcggc gggcagggcc gcgccctcgg ggcacgcgtc gcagcgcact 1800

ggctcttcgc ccgcgggcca cggctcggga gcgcaggggg ccgacggctc cgggacactg 1860

ggcagcgcgc tgggtgccga gggctccggg gccagggcag gggccgggcc ggggccggac 1920

ccggggcccg agccctggcg gagctgcagc agctcggaca gcgtgtggtt cttgcggagc 1980

tgaaggccgt cggggaaggg ctcctggcac agcgggcagc gggccgcgcc tccgggtccg 2040

ccggctccac tcgcgccacg atgcggccag agcgcgccca ggcaggcgag acagaagttg 2100

tggccgcagg gcagcgtcac cggctcccgg agtggctcta ggcagatggg gcagctgaag 2160

ggtccactgc cgtccatgac tccgcggccg cccagggcgc cgcggattgt gctccggcct 2220

gggagggacc cgggccgttc gcgccgcggc acctcccccc gggacctagg ccagggctcc 2280

-continued

ggcccccgcc cgcccccagc tcggcccgcc ccgcagcacc gcccgcctgc gggcccgcgg 2340

actcccagtc cccgcccaga caacgcaggg aggcctccga gcccgcgcga cccccagggg 2400

agtccgcgtg gtcctaagga agggcgctgc tgaaagggtg tgaaactgag tgagcgcggg 2460

cggagaacgc gagaggtaca gtggaaggaa tgggaaggac tcagcctcca tccctctggt 2520

ccttctgggc caggccagcg ctccgcaggg gctgggcgcg tctagctgtg tcatggctgc 2580

acctttgcgc tcgacccctc cctggcaccc ttagcaaagt tcctgagctg ctgactgcag 2640

agaaaaataa ctgctgatgt acatatgttt ttgcccctta gcgactcttt ttttttttt 2700

tttttttga gacagagttc ctctcttgtc gcccaggctg gagtgcaggg gcgctatctc 2760

ggctcactgc aacctccgcc tcccggattc aagtgattct cctgccttag cctcccaagt 2820

agctggaatt ataggcgctt gccaccacac caggcaaatt tttgtatttt tttgtagaga 2880

cggagtttca ccttgtccag gctggtcttg aactcgtgac ctcaggtgat ccacctgcct 2940

cggcctccca aagtgctggg attacaggcg tgagccacag cgcccagctc ctttagcgac 3000

tctttacccc aagtagggaa cttggacttt tctagacctg aaatgggaag ttgacgtcat 3060

tccctaagtc ccttgtctga ggaatgtagt tctggcccct ggacactggg gacacctgtg 3120

gtggggttgg gggttcccca actggaagta tttaaagctc atctaattat gcaaaaaaaa 3180

aaaaag 3186

<210> SEQ ID NO 34
<211> LENGTH: 2268
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 34

ggcggcgccc tgggcggccg cggagtcatg gacggcagtg gacccttcag ctgccccatc 60

tgcctagagc cactccggga gccggtgacg ctgccctgcg gccacaactt ctgtctcgcc 120

tgcctgggcg cgctctggcc gcatcgtggc gcgagtggag ccggcggacc cggaggcgcg 180

gcccgctgcc cgctgtgcca ggagcccttc cccgacggcc ttcagctccg caagaaccac 240

acgctgtccg agctgctgca gctccgccag ggctcgggcc ccgggtccgg ccccggcccg 300

gcccctgccc tggccccgga gccctcggca cccagcgcgc tgcccagtgt cccggagccg 360

tcggcccccc gcgctcccga gccgtggccc gcgggcgaag agccagtgcg ctgcgacgcg 420

tgccccgagg gcgcggccct gcccgccgcg ctgtcctgcc tctcctgcct cgcctccttt 480

tgccccgcgc acctgggccc gcacgagcgc agccccgccc tccgcggaca ccgcctggtg 540

ccgccgctgc gccggctaga ggagagcctg tgcccgcgcc acctacggcc gctcgagcgc 600

tactgccgcg cggagcgcgt gtgtctgtgc gaggcctgcg ccgcacagga gcaccgcggc 660

cacgagctgg tgccgctgga gcaggagcgc gcgcttcagg aggctgagca gtccaaagtc 720

ctgagcgccg tggaggaccg catggacgag ctgggtgctg gcattgcaca gtccaggcgc 780

acagtggccc tcatcaagag tgcagccgta gcagagcggg agagggtgag ccggctgttt 840

gcagatgctg cggccgccct gcagggcttc cagacccagg tgctgggctt catcgaggag 900

ggggaagctg ccatgctagg ccgctcccag ggtgacctgc ggcgacagga ggaacagcgc 960

agccgcctga gccgagcccg ccagaatctc agccaggtcc ctgaagctga ctcagtcagc 1020

ttcctgcagg agctgctggc actaaggctg gccctggagg atgggtgtgg ccctgggcct 1080

-continued

ggacccccga gggagctcag cttcaccaaa tcatcccaag ctgtccgtgc agtgagagac 1140

atgctggccg tggcctgcgt caaccagtgg gagcagctga gggggccggg tggcaacgag 1200

gatgggccac agaagctgga ctcggaagct gatgctgagc cccaagacct cgagagtacg 1260

aacctcttgg agagtgaagc tcccagggac tatttcctca agtttgccta tattgtggat 1320

ttggacagcg acacagcaga caagttcctg cagctgtttg gaaccaaagg tgtcaagagg 1380

gtgctgtgtc ctatcaacta ccccttgtcg cccacccgct tcacccattg tgagcaggtg 1440

ctgggcgagg gtgccctgga ccgaggcacc tactactggg aggtggagat tatcgagggc 1500

tgggtcagca tgggggtcat ggccgaagac ttctccccac aagagcccta cgaccgcggc 1560

cggctgggcc gcaacgccca ctcctgctgc ctgcagtgga atggacgcag cttctccgtc 1620

tggtttcatg ggctggaggc tcccctgccc caccccttct cgcccacggt tggggtctgc 1680

ctggaatacg ctgaccgtgc cttggccttc tatgctgtac gggacggcaa gatgagcctc 1740

ctgcggaggc tgaaggcctc ccggccccgc cggggtggca tcccggcctc ccccattgac 1800

cccttccaga gccgcctgga cagtcacttt gcggggctct tcacccacag actcaagcct 1860

gccttcttcc tggagagtgt ggacgcccac ttgcagatcg ggcccctcaa gaagtcctgc 1920

atatccgtgc tgaagaggag gtgatgccgg gcacgggcgc tcctgctgcc gtctctgctc 1980

caggaagctg cctcctctgg gccctctcct tcgtctggga aggcaccagc atgagtccca 2040

cacacccagc cttctcattt ctagaggctt ccacctttt atacactcag ccttccctct 2100

cccaggcagg aggaccccca gaccctgttc ccctgcagac ctcacttctg ggagacagag 2160

ctacagctgg gacagctcca agctacccta acccctcctt tcccaggttt ctagaatagt 2220

gtctggcatg tagtagatgc tcaataaaca cttgttgaat gaaaaaaa 2268

<210> SEQ ID NO 35
<211> LENGTH: 2113
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 35

gtatttttg ctaataaaat caagtatcaa attacttttg gtgagggatg ggcaactgta 60

ttgtactgct gataatttgc agtcaggaaa aattttgttt tccttctcat aacaacactg 120

ttaagccatg attcagttta ctttcacatt tttcttgagc caagcataga atgctattaa 180

attaaatttt attggtttta aattataatt ttagcctgtt tcaatctttg tgaaagctaa 240

ttctatcact caacataatt aagttattct cctcttgggc tcgtgccacc caaaatgatg 300

agctagcttt cgacaactga gtcttggtaa tttacaaaac tgttcagaaa gacgaaccaa 360

aatttgaatg gcttgctgtg ttctgtataa atttataatc gaaagtctga tcatggctac 420

aggttttttt ttttttaact tgttctgttt ttcaacaaac aatatactta tttgcaatta 480

aatgaaaata aagcattgaa ttctgaagtt ttagagaatg accttagaga tttttagtgg 540

ttgtcagcct tagcctccta tcaaaatcat ctgggaacta ttaaaagcat ttaatgttgt 600

gaccctacct agacctatta aattccattc tctggggtga agccaggcca ccagttattt 660

taaaagttct ccaagtgctt ctaatgtgca gcacaggttg ggagaatctg atctagctta 720

gcacttaaag gttatagtgc taccactaca taacacttct atattgagtt tactgtagta 780

tattaatgga gctatcttaa gtggaaaatg tctgatcaat cattaaatag caatctaagt 840

-continued

gaaatctata actaaaattt ctatctggaa agacaaaaag aaaaaccaac ttctctttga 900

gggtatgcta tacccagaga cttggctagt gctaaggagg gcatggtaga cttcccgagg 960

tcacacaaag aacaagcagc tcatgcttgg ttctacccaa tagtctgagt cccctcagtc 1020

actctgtttt agcctctaaa ccactaggac gtgatttgag aaatttcatg aagcctcaag 1080

gtgttctttc aggttatttg tacaatgggg aagaaaatat ctatagttca acctctttcc 1140

catactcatg acagatcacc tggcctgtgc ttcagtattt tacagattgt tcatctgtgg 1200

ctgaaaagct cttatcaaaa tgctcttact tttattgagc tgataaagtt gccatcctaa 1260

cttctatcca ttggttctgg ctctgcctcc tggaggaaca cagactatgt tgattgtctt 1320

ttccctgtgt tagcaatgca gatggattgt aattttccta aatctgctct tcaagaaaag 1380

atttcaagtt ttgaaaacca tatcatcatc ctgtcctcgg tgacacctgt ttctctcctt 1440

tgtcaacatc acctttcaca atgtgtcacc tggatagcag atcgaacaac ctggctataa 1500

tagtggttct gatagcctcc tttgatctcg tcattgtatt ctagaaatga agcccaagat 1560

tttatgtatt tttaccgcgg tcataccacg ttgatttaaa agatcaatga ttttcgagcc 1620

aggctggaca ttaaaatcac ttgggagagc tttgaaaaaa tatagcgata tcctgtttta 1680

cccttgatca attaaatcag aatctctggg gataaggccc aggtacagtg atttttaaaa 1740

gtacttcaat gaccctaata tgcatctaag tgagttctat ccttaaatgt ggtttctctc 1800

aatttgtagg taaatacaat attttgttgt tgtttgaatg taaatatcag gcttttgatc 1860

tgtcctagtt taatttttctg ataatagtag tcccagaatt ttcaaatgca attctttaag 1920

attgcaatga agttagaact cttacgttac tgtttatgtc cacaaaatgt ttagaaaagt 1980

ctgaaaagga ttttgaaaaa gattaacagt gttaatattt cttatattgt tgggtttatg 2040

cataatataa tttggttaat atgtatcata aataaaagta ttttaagatc acaaaaagaa 2100

aaaaaaaaaa aac 2113

<210> SEQ ID NO 36
<211> LENGTH: 4072
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 36

acaccgccct cccgccagac tcccggcggc tcctcctccc tctcccaaac ccactcccaa 60

agctaagtgc aggcttcccc gttccagcca gaagcgctgc gtgagcctcc acacgtagcc 120

gcaggcagct ccttaaatag cgtccgcgct gagcaaacag tccagacgtg gggcccagga 180

gggcgagctg aggcgaccgc accgggcgcg cagcggcggc gggtcagccg gcggccaata 240

gccagggcgc ggcccgcccc gtcgcctccc ctcggggagc ctataaggcc tccgcagcgc 300

cccgggcgcc tgctgctccg tgcctccacc gacgacctca ctcagctgcg ttacgcgccg 360

ctccggctgc cggccgcgcg ccttgcccgc cggctcccgc ccgcaatcgg cggctcaggg 420

cggacccggg tctctgcgtt ctcgcgagaa gcgcggcgct gcggggccgt gggcgcctga 480

gcccgcgcgg ccctcgaggg ccgaatatgg ggggatgcac ggtgaagcct cagctgctgc 540

tcctggcgct cgtcctccac ccctggaatc cctgtctggg tgcggactcg gagaagccct 600

cgagcatccc cacagataaa ttattagtca taactgtagc aacaaaagaa agtgatggat 660

tccatcgatt tatgcagtca gccaaatatt tcaattatac tgtgaaggtc cttggtcaag 720

-continued

gagaagaatg gagaggtggt gatggaatta atagtattgg agggggccag aaagtgagat 780

taatgaaaga agtcatggaa cactatgctg atcaagatga tctggttgtc atgtttactg 840

aatgctttga tgtcatattt gctggtggtc cagaagaagt tctaaaaaaa ttccaaaagg 900

caaaccacaa agtggtcttt gcagcagatg gaattttgtg gccagataaa agactagcag 960

acaagtatcc tgttgtgcac attgggaaac gctatctgaa ttcaggagga tttattggct 1020

atgctccata tgtcaaccgt atagttcaac aatggaatct ccaggataat gatgatgatc 1080

agctctttta cactaaagtt tacattgatc cactgaaaag ggaagctatt aacatcacat 1140

tggatcacaa atgcaaaatt ttccagacct taaatggagc tgtagatgaa gttgttttaa 1200

aatttgaaaa tggcaaagcc agagctaaga atacatttta tgaaacatta ccagtggcaa 1260

ttaatggaaa tggacccacc aagattctcc tgaattattt tggaaactat gtacccaatt 1320

catggacaca ggataatggc tgcactcttt gtgaattcga tacagtcgac ttgtctgcag 1380

tagatgtcca tccaaacgta tcaataggtg tttttattga gcaaccaacc ccttttctac 1440

ctcggtttct ggacatattg ttgacactgg attacccaaa agaagcactt aaacttttta 1500

ttcataacaa agaagtttat catgaaaagg acatcaaggt atttttttgat aaagctaagc 1560

atgaaatcaa aactataaaa atagtaggac cagaagaaaa tctaagtcaa gcggaagcca 1620

gaaacatggg aatggacttt tgccgtcagg atgaaaagtg tgattattac tttagtgtgg 1680

atgcagatgt tgttttgaca aatccaagga ctttaaaaat tttgattgaa caaaacagaa 1740

agatcattgc tcctcttgta actcgtcatg gaaagctgtg gtccaatttc tggggagcat 1800

tgagtcctga tggatactat gcacgatctg aagattatgt ggatattgtt caagggaata 1860

gagtaggagt atggaatgtc ccatatatgg ctaatgtgta cttaattaaa ggaaagacac 1920

tccgatcaga gatgaatgaa aggaactatt ttgttcgtga taaactggat cctgatatgg 1980

ctctttgccg aaatgctaga gaaatgactt tacaaaggga aaaagactcc cctactccgg 2040

aaacattcca aatgctcagc cccccaaagg gtgtatttat gtacatttct aatagacatg 2100

aatttggaag gctattatcc actgctaatt acaatacttc ccattataac aatgacctct 2160

ggcagatttt tgaaaatcct gtggactgga aggaaaagta tataaaccgt gattattcaa 2220

agattttcac tgaaaatata gttgaacagc cctgtccaga tgtcttttgg ttccccatat 2280

tttctgaaaa agcctgtgat gaattggtag aagaaatgga acattacggc aaatggtctg 2340

ggggaaaaca tcatgatagc cgtatatctg gtggttatga aaatgtccca actgatgata 2400

tccacatgaa gcaagttgat ctggagaatg tatggcttca ttttatccgg gagttcattg 2460

caccagttac actgaaggtc tttgcaggct attatacgaa gggatttgca ctactgaatt 2520

ttgtagtaaa atactcccct gaacgacagc gttctcttcg tcctcatcat gatgcttcta 2580

catttaccat aaacattgca cttaataacg tgggagaaga ctttcaggga ggtggttgca 2640

aatttctaag gtacaattgc tctattgagt caccacgaaa aggctggagc ttcatgcatc 2700

ctgggagact cacacatttg catgaaggac ttcctgttaa aaatggaaca agatacattg 2760

cagtgtcatt tatagatccc taagttattt acttttcatt gaattgaaat ttattttgga 2820

tgaatgactg gcatgaacac gtctttgaag ttgtggctga gaagatgaga ggaatattta 2880

aataacatca acagaacaac ttcactttgg gccaaacatt tgaaaaactt tttataaaaa 2940

attgtttgat atttcttaat gtctgctctg agccttaaaa cacagattga agaagaaaag 3000

-continued

| | |
|---|---|
| aaagaaaaaa cttaaatatt tatttctatg ctttgttgcc tctgagaata atgacaattt | 3060 |
| atgaatttgt gtttcaaatt gataaaatat ttaggtacaa ataacaagac taataatatt | 3120 |
| ttcttattta aaaaaagcat gggaagattt ttatttatca aaatatagag gaaatgtaga | 3180 |
| caaaatggat ataaatgaaa attaccatgt tgtaaaacct tgaaaatcag attctaactg | 3240 |
| gatttgtatg caactaagta tttttctgaa cacctatgca ggtcttattt acagtagtta | 3300 |
| ctaagggaac acacaaagaa ttacacaacg ttttcctcaa gaaaatggta caaaacacaa | 3360 |
| ccgaggagcg tatacagttg aaaacatttt tgttttgatt ggaaggcaga ttattttata | 3420 |
| ttagtattaa aaatcaaacc ctatgtttct ttcagatgaa tcttccaaag tggattatat | 3480 |
| taagcaggta ttagatttag gaaaaccttt ccatttctta aagtattatc aagtgtcaag | 3540 |
| atcagcaagt gtccttaagt caaacaggtt ttttttgttg ttgtttttgc tttgtttcct | 3600 |
| tttttagaaa gttctagaaa ataggaaaac gaaaaatttc attgagatga gtagtgcatt | 3660 |
| taattatttt ttaaaaaact ttttaagtac ttgaatttta tatcaggaaa acaaagttgt | 3720 |
| tgagccttgc ttcttccgtt ttgccctttg tctcgctcct tattcttttt ttgggggggag | 3780 |
| ggttatttgc ttttttatct tcctggcata atttccattt tattcttctg agtgtctatg | 3840 |
| ttaacttccc tctatcccgc ttataaaaaa attctccaac aaaaatactt gttgacttga | 3900 |
| tgttttatca cttctctaag taaggttgaa atatccttat tgtagctact gtttttaatg | 3960 |
| taaaggttaa acttgaaaag aaattcttaa tcacggtgcc aaaattcatt ttctaacacc | 4020 |
| atgtgttaga aaattataaa aaataaaata attttagaaa aaaaaaaaaa aa | 4072 |

What is claimed is:

**1**. A method for identifying putative therapeutic targets comprising:

Obtaining a paired brain tissue sample from a live human wherein each member of the pair has a different level of electrical brain activity from the other member;

Identifying long non-protein-coding RNA (lncRNA) molecules (lncRNAs) and protein-coding messenger RNA (mRNA) molecules (mRNAs) that are differentially expressed between the members of each individual sample pair;

Linking a first differentially expressed lncRNA with a differentially expressed mRNA and/or a second differentially expressed lncRNA when the gene encoding the first differentially expressed lncRNA overlaps with, or is adjacent to, the gene encoding the differentially expressed mRNA and/or the gene encoding the differentially expressed second lncRNA along the human genome, thereby identifying an lncRNA/mRNA gene pair and/or an lncRNA/lncRNA gene pair as putative cis-encoded therapeutic targets; and/or

Linking a first differentially expressed lncRNA with a differentially expressed mRNA and/or with a second differentially expressed lncRNA when the differentially expressed first lncRNA and the differentially expressed mRNA and/or second lncRNA are encoded at different genomic loci, thereby identifying an lncRNA/mRNA gene pair and/or an lncRNA/lncRNA gene pair as putative trans-encoded therapeutic targets.

**2**. A method of claim **1** wherein linking of differentially expressed lncRNA with differentially expressed mRNA and/or lncRNA further requires that the differential expression of the lncRNA and mRNA or lncRNA and lncRNA be observed in more than one brain sample pair, each pair having a low electrical brain activity member and a high electrical brain activity member.

**3**. A method of claim **1** or **2** wherein electrical brain activity is classified as high or low based on the frequency and/or amplitude of interictal and ictal spiking.

**4**. A method of claim **1**, **2** or **3** wherein differential expression is identified by quantifying lncRNA and mRNA expression.

**5**. A method of claim **4** wherein the expression quantification utilizes at least one microarray capable of quantifying lncRNA expression and mRNA expression.

**6**. A method of claim **4** or **5** wherein the quantifying utilizes at least one microarray capable of quantifying lncRNA expression and at least one microarray capable of quantifying mRNA expression wherein consistency of differential expression data between the at least one lncRNA microarray and the at least one mRNA microarray is evaluated by correlating the fold-change of protein-coding control genes common to both arrays.

**7**. A method of claim **4**, **5** or **6** further comprising evaluating the putative therapeutic target as a molecular site of effective intervention.

**8**. A method of claim **7** wherein the therapeutic target of a pair is lncRNA, mRNA and/or both.

**9**. A microarray for identifying putative therapeutic targets in the human brain comprising probes for lncRNA and probes for mRNA wherein at least a subset of the mRNA probes is included based on the representation of their corresponding genes by probes on a different genomewide expression analysis microarray.

**10**. A microarray of claim **8** or **9** wherein the lncRNA probes are 50-mer to 70-mer probes mapped to a single genomic location.

**11**. A microarray of claim **8**, **9** or **10** wherein the lncRNA probes are free of interspersed and simple repeats and segmental duplications.

**12**. A microarray of claim **8**, **9**, **10** or **11** comprising 7 or 8 distinct probes per lncRNA.

**13**. A microarray of claim **8**, **9**, **10**, **11** or **12** comprising probes for at least 1000 lncRNA genes.

**14**. A method of assessing putative therapeutic targets in the human brain comprising:

Exposing human neuroblastoma cells to either a single depolarization or repeated depolarizations;

Identifying time-dependent differential lncRNA and mRNA expression in the cells exposed to either single and/or repeated depolarizations, relative to untreated control cells; and

(i) Linking a first differentially expressed lncRNA with differentially expressed mRNAs and/or second differentially expressed lncRNAs when the gene encoding the first differentially expressed lncRNA overlaps with, or is adjacent to the gene encoding the differentially expressed mRNA or second differentially expressed lncRNA thereby identifying lncRNA/mRNA and/or lncRNA/lncRNA gene pairs as putative cis-encoded therapeutic targets; and/or

(ii) Linking a first differentially expressed lncRNA with a differentially expressed mRNA and/or with a second differentially expressed lncRNA when the first lncRNA and mRNA and/or second lncRNA are encoded at different genomic loci, thereby identifying lncRNA/mRNA and/or lncRNA/lncRNA gene pairs as putative trans-encoded therapeutic targets.

**15**. A method comprising targeting the first lncRNA of an lncRNA/mRNA or lncRNA/lncRNA gene pair as a putative therapeutic target wherein the first lncRNA and mRNA or second lncRNA are differentially expressed in areas of the brain having a different characteristic demonstrated to be relevant in one or more of epileptic activity, inflammation, cellular proliferation multiple sclerosis, neurodegeneration or autism and/or have been linked because the gene encoding the differentially expressed lncRNA overlaps with, or is adjacent to the gene encoding the differentially expressed mRNA or second differentially expressed lncRNA.

**16**. A method of claim **15** wherein the gene pair is BDNFOS (SEQ ID NO: 1)/BDNF (SEQ ID NO: 2); AF086035 (SEQ ID NO: 3)/MAPK1IP1L (SEQ ID NO: 4); AK093366 (SEQ ID NO: 5)/AG2 (SEQ ID NO: 6); BC047792 (SEQ ID NO: 7)/PURB (SEQ ID NO: 8); AK096235 (SEQ ID NO: 9)/LCP1 (SEQ ID NO: 10); AL110130 (SEQ ID NO: 11)/SMEK2 (SEQ ID NO: 12); BC013641 (SEQ ID NO: 13)/ARC (SEQ ID NO: 14); hTF27297 (SEQ ID NO: 15)/CYR61 (SEQ ID NO: 16); RPPH1 (SEQ ID NO: 17)/NEAT1 (SEQ ID NO: 18); NEAT1 (SEQ ID NO: 18)/EGR3 (SEQ ID NO: 19); NEAT1 (SEQ ID NO: 18)/CR600638 (SEQ ID NO: 20); RPPH1 (SEQ ID NO: 17)/NEAT2 (MALAT1) (SEQ ID NO: 21); CR600638 (SEQ ID NO: 20)/FLT1 (SEQ ID NO: 22); BC012463 (SEQ ID NO: 23)/LRRN1 (SEQ ID NO: 24); CR615000 (SEQ ID NO: 25)/ERAP1 (SEQ ID NO: 26); BCO28229 (SEQ ID NO: 27)/BTG3 (SEQ ID NO: 28); BC078172 (SEQ ID NO: 29)/VIM (SEQ ID NO: 30); AF075087 (SEQ ID NO: 31)/S1PR1 (SEQ ID NO: 32); AK123944 (SEQ ID NO: 33)/TRIM47 (SEQ ID NO: 34); or AL110176 (SEQ ID NO: 35)/PLOD2 (SEQ ID NO: 36).

**17**. A method comprising targeting an lncRNA gene as a site of putative therapeutic intervention wherein the lncRNA gene is differentially expressed in at least one area of the brain having a different characteristic then a second area and wherein the lncRNA gene is KCNQ1OT1; RPPH1; NEAT1; NEAT2 or MIAT.

**18**. A method comprising targeting a gene as a site of therapeutic intervention when the gene was identified by a method or microarray of any one of claims **1**-**17**.

**19**. A method of claim **18** wherein the targeting includes gene silencing or gene activating.

**20**. A method of claim **18** or **19** wherein the targeting includes gene silencing through RNA interference.

* * * * *