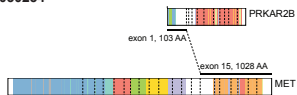


Figure S3.

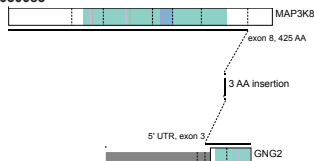
SJWLM030234



PRKAR2B NM\_002736

- DD\_Ribbeta\_PKA Dimerization/Docking domain of the Type II beta Regulatory subunit of cAMP-de...
- other dimer interface [polypeptide binding]
- other AKAP interaction site [polypeptide binding]
- Crp cAMP-binding domain of CRP or a regulatory subunit of cAMP-dependent protein ...
- CAP\_ED effector domain of the CAP family of transcription factors; members include C...
- other ligand binding site [chemical binding]
- other flexible hinge region
- MET NM\_00127500
- Sema\_MET The Sema domain, a protein interacting module, of MET (also called hepatocyte...
- other HGF binding site [polypeptide binding]
- other InB binding site [polypeptide binding]
- PSI Domain found in Plexins, Semaphorins and Integrins
- IPT\_plexin\_repeat1 First repeat of the IPT domain of Plexins and Cell Surface Receptors (PCSR) ...
- IPT\_plexin\_repeat2 Second repeat of the IPT domain of Plexins and Cell Surface Receptors (PCSR) ...
- IPT Immunoglobulin-like fold, Plexins, Transcription factors (IPT). IPTs are also...
- IPT Ig-like, plexins, transcription factors
- PKinase\_Tyr Protein tyrosine kinase
- PTKc\_Met\_Ron Catalytic domain of the Protein Tyrosine Kinases, Met and Ron
- other ATP binding site [chemical binding]
- other polypeptide substrate binding site [polypeptide binding]
- other activation loop (A-loop)

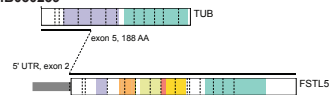
SJMEL030083



MAP3K8 NM\_005204

- STKc\_MAP3K8 Catalytic domain of the Serine/Threonine kinase, Mitogen-Activated Protein Ki...
- other ATP binding site [chemical binding]
- other polypeptide substrate binding site [polypeptide binding]
- other activation loop (A-loop)
- GNG2 NM\_063064
- GGL G protein gamma subunit-like motifs
- other beta subunit binding site [polypeptide binding]

SJMB030239



TUB NM\_003320

- Tub\_N Tubby N-terminal
- Tub Tub family
- FSTL5 NM\_020116
- KAZAL\_FS Kazal type serine protease inhibitors and follistatin-like domains. Kazal inh...
- EFh EF-hand, calcium binding motif. A diverse superfamily of calcium sensors and ...
- EF-hand\_7 EF-hand domain pair
- other Ca2+ binding site [ion binding]
- G\_like Immunoglobulin like
- Gc2 Immunoglobulin C-2 type
- set Immunoglobulin I-set domain
- Ig2\_Follistatin\_like Second immunoglobulin (Ig)-like domain of a follistatin-like molecule encoded...
- YncE DNA-binding beta-propeller fold protein YncE [General function prediction only]

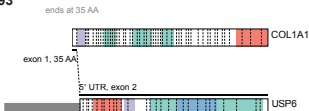
SJHGG030166



CLIP2 NM\_003388

- CAP\_GLY CAP-Gly domain
- Smc Chromosome segregation ATPase (Cell cycle control, cell division, chromosome ...
- DUF4349 Domain of unknown function (DUF4349)
- ALK NM\_004304
- MAM MAM domain, meprin/A5mu
- MAM Meprin, A5 protein, and protein tyrosine phosphatase Mu (MAM) domain. MAM is ...
- LDLA Low-density lipoprotein receptor domain class A
- Fka\_inhibition Coagulation Factor Xa inhibitory site
- PTKc\_Alk\_Ltk Catalytic domain of the Protein Tyrosine Kinases, Anaplastic Lymphoma Kinase ...
- PKinase\_Tyr Protein tyrosine kinase
- other ATP binding site [chemical binding]
- other polypeptide substrate binding site [polypeptide binding]
- other activation loop (A-loop)

SJST030093

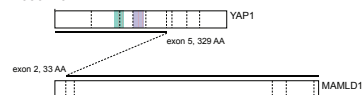


starts at 1688 bp in 5' UTR

COL1A1 NM\_000088

- VVC von Willebrand factor type C domain
- Collagen Collagen triple helix repeat (20 copies)
- COLFI Fibrillar collagen C-terminal domain
- USP6 NM\_044505
- TBC Domain in Tre-2, Bub2p, and Cdc16p. Probable Rab-GAPs
- DUF4607 Domain of unknown function (DUF4607)
- UBP12 Ubiquitin C-terminal hydrolase [Posttranslational modification, protein turno...
- Peptidase\_C19 Peptidase C19 contains ubiquitiny hydrolyases. They are intracellular peptid...

SJEPD030146



YAP1 NM\_001130145

- WW Two conserved tryptophans domain; also known as the WWP or rsp5 domain; aroun...
- binding binding pocket
- WW Domain with 2 conserved Trp (W) residues

SJDSRCT030041



MOB2 NM\_001122233

- Mob1\_phocain Mob1/phocain family
- TERT NM\_001193376
- Telomerase\_RBD Telomerase ribonucleoprotein complex - RNA binding domain
- RT\_like RT\_like: Reverse transcriptase (RT, RNA-dependent DNA polymerase)\_like family...
- TERT TERT: Telomerase reverse transcriptase (TERT). Telomerase is a ribonucleoprot...
- other putative nucleic acid binding site [nucleotide binding]

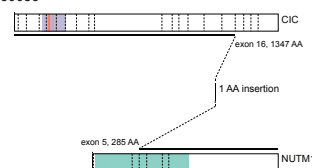
SJDSRCT030041



EWSR1 NM\_01163285

- RRM RNA recognition motif (RRM) domain [Translation, ribosomal structure and biog...
- RRM\_EWS RNA recognition motif in vertebrate Ewing Sarcoma Protein (EWS)
- zf-RanBP Zn-finger in Ran binding protein and others
- RanBP2-type Zn finger RanBP2-type Zn finger [structural motif]
- other Zn binding site [ion binding]
- WT1 NM\_001198552
- WT1 Wilms' tumour protein
- COG5048 FOG: Zn-finger [General function prediction only]
- C2H2 Zn finger C2H2 Zn finger [structural motif]
- other putative nucleic acid binding site [nucleotide binding]
- zf-H2C2\_2 Zinc-finger double domain
- other Zn binding site [ion binding]

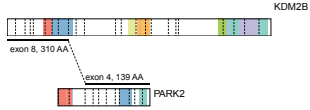
SJBT030086



CIC NM\_015125

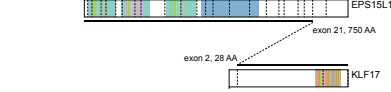
- NHP6B Chromatin-associated proteins containing the HMG domain [Chromatin structure ...
- SOX.TCF\_HMG-box SOX.TCF\_HMG-box, class I member of the HMG-box superfamily of DNA-binding pro...
- DNA binding DNA binding site [nucleotide binding]
- NUTM1 NM\_175741
- NUT NUT protein

**SJST030084**



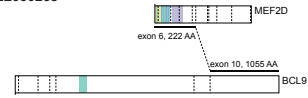
- KDM2B NM\_032890**
- JmyC A domain family that is part of the cupin metalloenzyme superfamily
  - cupin\_like Conserved domain found in cupin and related proteins
  - zf-CXXC CXXC zinc finger domain
  - PHD\_KDM2B PHD finger found in Lysine-specific demethylase 2B (KDM2B)
  - other putative histone H3 binding site [polypeptide binding]
  - other Zn binding site [ion binding]
  - F-box-like F-box like
  - leucine-rich repeat leucine-rich repeat [structural motif]
  - AMN1 Antagonist of mitotic exit network protein 1
- PARK2 NM\_004862**
- UBQ Ubiquitin homologues
  - parkin\_N amino-terminal ubiquitin-like of parkin protein
  - other Arg42Pro mutation site
  - IBR In Between Ring fingers
  - IBR IBR domain, a half RING-finger domain

**SJBT030076**



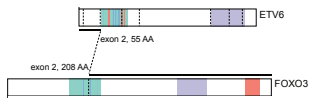
- EPS15L1 NM\_021235**
- EH Eps15 homology domain
  - EH Eps15 homology domain, found in proteins implicated in endocytosis, vesicle...
  - other pseudo EF-hand loop
  - other peptide binding pocket
  - other Ca2+ binding site [ion binding]
  - Leberleini Ciliary protein causing Leber congenital amaurosis disease
  - SPEC Spectrin repeats, found in several proteins involved in cytoskeletal structur...
- KLF17 NM\_173484**
- COG6848 FOG, Zn-finger [General function prediction only]
  - zf-C2H2 Zinc finger, C2H2 type
  - C2H2 Zn finger C2H2 Zn finger [structural motif]
  - other Zn binding site [ion binding]
  - other putative nucleic acid binding site [nucleotide binding]
  - zf-H2C2\_2 Zinc-finger double domain

**SJBALL030268**



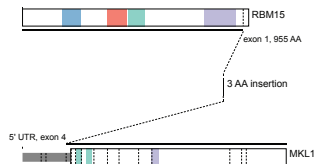
- MEF2D NM\_009320**
- MADS\_MEF2\_like MEF2 (myocyte enhancer factor 2)-like/Type II subfamily of MADS (MCM1, Agamo...
  - DNA binding DNA binding site [nucleotide binding]
  - other dimerization interface [polypeptide binding]
  - other putative phosphorylation site [posttranslational modification]
  - HJURP\_C Holiday junction regulator protein family C-terminal repeat
- BCL9 NM\_004326**
- BCL9 B-cell lymphoma 9 protein

**SJBALL030052**



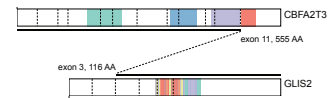
- ETV6 NM\_001987**
- SAM\_PNT-Tel\_Yan Sterile alpha motif (SAM)/Pointed domain of Tel/Yan protein
  - other oligomer interface ML [polypeptide binding]
  - other oligomer interface EH [polypeptide binding]
  - ETS erythroblast transformation specific domain
- FOXO3 NM\_001455**
- FKH Forkhead (FHL), also known as a 'winged helix'. FHL is named for the Drosophila...
  - DNA binding DNA binding site [nucleotide binding]
  - FOXO\_KIX\_bdg KIX-binding domain of forkhead box O, CR2
  - FOXO-TAD Transactivation domain of FOXO protein family

**SJAML030209**



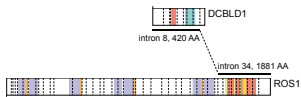
- RBM15 NM\_022768**
- RBM1\_RBM15 RNA recognition motif 1 in vertebrate RNA binding motif protein 15 (RBM15)
  - RBM2\_RBM15 RNA recognition motif 2 in vertebrate RNA binding motif protein 15 (RBM15)
  - RBM3\_RBM15 RNA recognition motif 3 in vertebrate RNA binding motif protein 15 (RBM15)
  - SPOC SPOC domain
- MKL1 NM\_020831**
- RPEL Repeat in Drosophila CG10860, human KIAA0680 and C. elegans F20H9.2
  - SAP SAP domain

**SJAML030120**



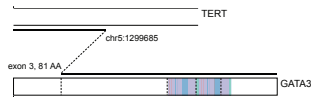
- CBFA2T3 NM\_005187**
- TAFH NHR1 homology to TAF
  - NHR2 NHR2 domain like
  - vATP-synt\_E ATP synthase (E/F1 kDa) subunit
  - zf-MYND MYND finger
- GLIS2 NM\_032575**
- C2H2 Zn finger C2H2 Zn finger [structural motif]
  - other putative nucleic acid binding site [nucleotide binding]
  - zf-H2C2\_2 Zinc-finger double domain
  - COG6848 FOG, Zn-finger [General function prediction only]
  - other Zn binding site [ion binding]

**SJTAL030249**



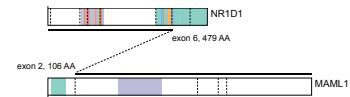
- GOPC NM\_020399**
- ZIP\_2 Basic region leucine zipper
  - other dimer interface [polypeptide binding]
  - PDZ\_signaling PDZ domain found in a variety of Eumetazoa signaling molecules, often in tan...
  - other protein binding site [polypeptide binding]
- ROS1 NM\_002944**
- FN3 Fibronectin type 3 domain. One of three types of internal repeats found in th...
  - other Cytokine receptor motif
  - other Interdomain contacts
  - Pkinase\_Tyr Protein tyrosine kinase
  - PTK\_c\_ross Catalytic domain of the Protein Tyrosine Kinase, C-ros
  - other ATP binding site [chemical binding]
  - other polypeptide substrate binding site [polypeptide binding]
  - other activation loop (A-loop)

**SJNBL030147**



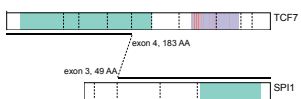
- GATA3 NM\_001002295**
- ZnF\_GATA zinc finger binding to DNA consensus sequence [AT]GATA[GAC]
  - ZnF\_GATA Zinc finger DNA binding domain, binds specifically to DNA consensus sequence...
  - other zinc binding site [ion binding]
  - DNA binding DNA-binding region [nucleotide binding]

**SJST030122**



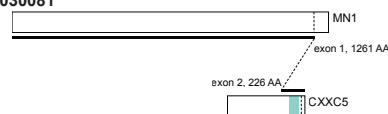
- NR1D1 NM\_021724**
- NR\_DBD\_REV\_ERB DNA-binding domain of REV-ERB receptor like is composed of two C4-type zinc f...
  - other zinc binding site [ion binding]
  - DNA binding DNA binding site [nucleotide binding]
  - other dimer interface [polypeptide binding]
  - NR\_LBD The ligand binding domain of nuclear receptors, a family of ligand-activated...
  - other ligand binding site [chemical binding]
  - other coregulator recognition site
- MAML1 NM\_014757**
- MAML-1 Maml-1 domain
  - TFlIA Transcription factor IIA, alpha/beta subunit

**SJTAL030263**



- TCF7 NM\_003202**
- CTNMB1\_binding N-terminal CTNMB1 binding
  - SOX-TCF\_HMG-box SOX-TCF\_HMG-box, class I member of the HMG-box superfamily of DNA-binding pro...
  - DNA binding DNA binding site [nucleotide binding]
- SP11 NM\_001080547**
- ETS erythroblast transformation specific domain

**SJBT030081**



- CXXC5 NM\_016463**
- zf-CXXC CXXC zinc finger domain

**Figure S3.** Schematic depiction of breakpoints encoding rare fusion proteins. The 20 rare fusions identified in the G4K cohort are depicted as protein structures with exon boundaries indicated by vertical dotted lines. Functional domains of each protein are color coded with the color key shown below each fusion pair. The portions of each protein found in the final fusion product is indicated by a solid line with exons involved in the fusion junction labeled and connected by a dotted line.