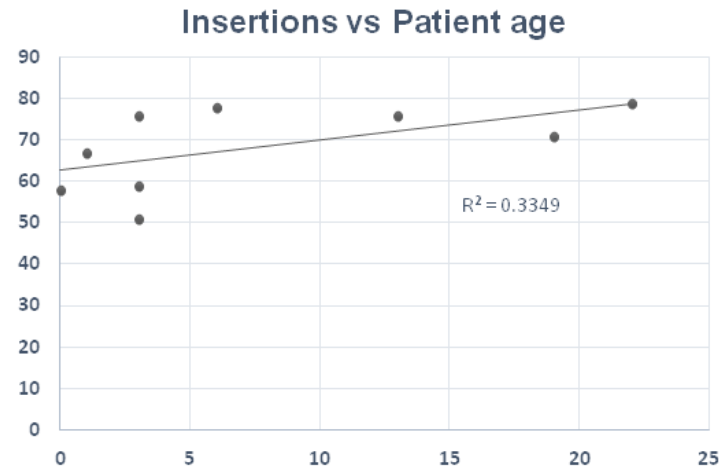


**SUPPORTING INFORMATION FOR THE ARTICLE:**

**Somatically acquired LINE-1 insertions in normal esophagus undergo clonal expansion in esophageal squamous cell carcinoma**

Somatic insertion occurrence, ORF1p expression, and patient age			
Sample designation	Age	insertions	Expression
1N/T	76	3	N/A
2N/T	67	1	N/A
3N/T	51	3	N/A
8N/T	78	6	N/A
20N/T	79	22	+++
21N/T	71	19	+++
22N/T	59	3	N/A
23N/T	76	13	+++
24N/T	58	0	+



**Figure S1: Somatic insertion occurrence, ORF1p expression, and patient age**

**Figure S1: Somatic insertion occurrence, ORF1p expression, and patient age.**

This figure details the ages of the patients compared with the number of somatic insertions which occurred in each individual and the level of ORF1p expression observed in the individual. The graph shows there is a small positive correlation between the age of the patients and the insertion occurrence into them, e.g. older patients tend to have more somatic insertions. However, the positive correlation between age and insertion occurrence is not statistically significant when a linear regression is performed ( $P = 0.1025$ ).

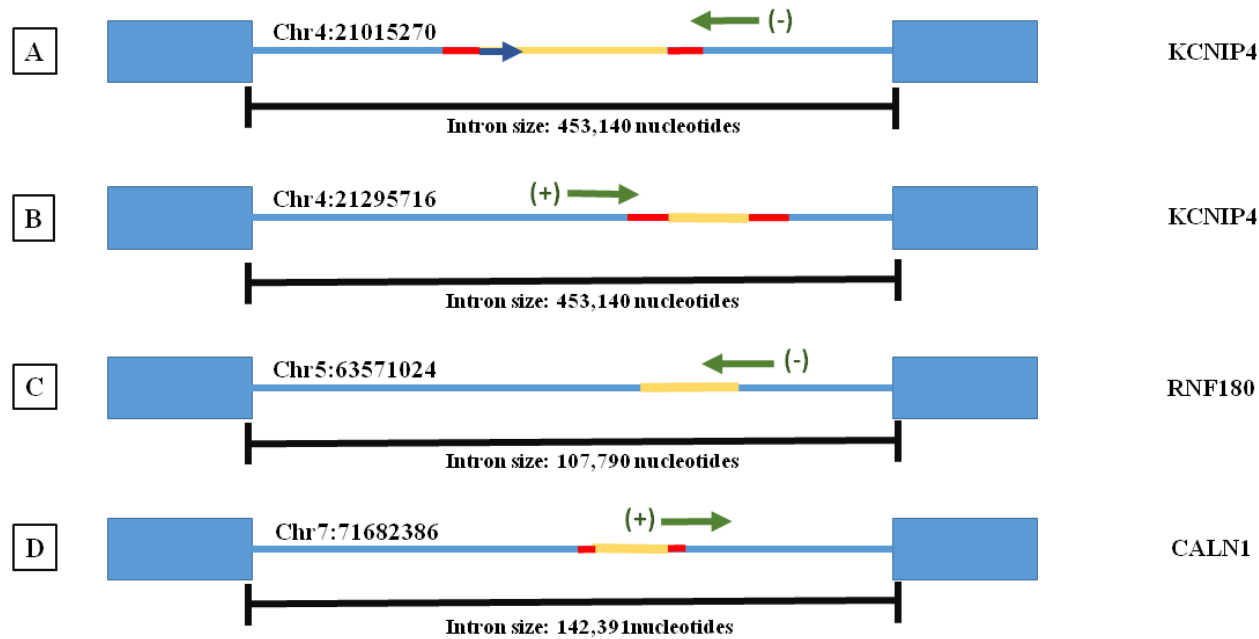


Figure S2: Examples of Somatic Insertions into Genes

**Figure S2: Examples of Somatic Insertions into Genes**

Several somatic insertions into the introns of genes are diagrammed in this figure along with the introns sizes, and where applicable the hallmarks of target-primed reverse transcription. (A) An insertion into KCNIP4 is pictured here in yellow between two exons drawn in blue. The green arrow indicates the insertion occurred into the antisense strand and the red lines depict the target site duplications. The dark blue arrow indicates that the insertion had a 5' inversion of its sequence. (B) A second insertion into the KCNIP4 gene is

pictured here in yellow flanked by target site duplications in red. The green arrow indicates the insertion occurred into the sense strand. (C) An insertion into RNF180 is shown in yellow without target site duplications; it is presumed to be an endonuclease independent insertion. The insertion occurred into the antisense strand as indicated by the green arrow. (D) An insertion into CALN1 is depicted by the yellow line and there are target site duplications on either side in red. The insertion occurred into the sense strand.

**Table S1: Genes previously associated with cancer**

Gene Name(s)	Cancer type(s)	NGC 4.0	Associated with smoking?	Cancer Reference	Smoking Reference	OMIM	Entrez Gene
EYS	esophageal, lung	Y	Y	(Dulak, et al., 2013; Kan, et al., 2010)	(Eckel-Passow, et al., 2014; Rose, et al., 2010)	612424	346007
PCDH15	head and neck, pancreatic	Y	Y	(Jones, et al., 2008; Stransky, et al., 2010)	(Rose, et al., 2010; Uhl, et al., 2008)	605514	65217
HDAC9	adrenocortical adenoma, squamous cell carcinoma	Y	Y	(Cao, et al., 2014; Fleming, et al., 2010)	(Rose, et al., 2010)	606543	9734
PTPRD	lung, breast, pancreas, gastric adenocarcinoma	Y	Y	(Ding, et al., 2008; Koboldt, et al., 2012)	(Rose, et al., 2010; Uhl, et al., 2008)	601598	5789
KCNIP4	renal cell carcinoma, haematological	N	Y	(Bonne, et al., 2007; Lauc G, 2013)	(Rose, et al., 2010; Tang, et al., 2010)	608182	80333
RNF180	gastric	N	Y	(Cheung, et al., 2012; Deng, et al., 2014)		616015	285671
PPP2R2B	breast, colorectal	N	Y	(Tan, et al., 2010; Vazquez, et al., 2010)	(Dong, et al., 2012; Rose, et al., 2010)	604325	5521
TRMT1	prostate	N		(Wen, et al., 2014)		611669	55621
RPS6KA2	ovarian, colorectal, breast, pancreatic	N	Y	(Bignone, et al., 2007; Milosevic, et al., 2010)	(Rose, et al., 2010; Tsaprouni, et al., 2010)	601685	6196
C8ORF37-AS1	colorectal	N		(Fernandez-Rozadilla, et al., 2013)		614477	100616530
MOB3B	prostate	N	Y	(Kerns, et al., 2013)	(Rose, et al., 2010; Uhl, et al., 2008)		79817
CELF2	breast, colon, squamous cell carcinoma	N	Y	(Matta, et al., 2009; Mukhopadhyay, et al., 2010)	(Rose, et al., 2010)	602538	10659
MAGI2	lung adenocarcinoma	N		(Kitamura, et al., 2014)		606382	9863
PTPRM	breast, glioblastoma, B cell leukemia	N	Y	(Burgoyne, et al., 2009; Sun, et al., 2010)	(Rose, et al., 2010)	176888	5797
BPIFA2	salivary gland tumors	N		(Gonzalez-Arriagada, et al., 2015; Vargas, et al., 2008)			140683
SELP	colon carcinoma metastasis, colorectal cancer, colon carcinoma	N		(Dymicka-Piekarska and Kemona, 2009; Kohler, et al., 2010; Wei, et al., 2010)		173610	6403
AUTS2	haematological	N	Y	(Lauc G, 2013)	(Rose, et al., 2010)		26053
MORC1	colorectal cancer	N	Y	(Figueiredo, et al., 2014)	(Rose, et al., 2010)	603205	27136
NLGN1	oral squamous cell carcinoma, uterine leiomyosarcoma	N	Y	(Cha, et al., 2011; Davidson, et al., 2010)	(Rose, et al., 2010)	600568	22871
FARS2	colon, colorectal	N	Y	(Won, et al., 2012)	(Rose, et al., 2010)	611592	10667
CALN1	gastric	N	Y	(Li, et al., 2014)	(Rose, et al., 2010)	607176	83698
TENM4/ODZ4/DOC4	primary lymphoma of CNS, neuroblastoma, breast cancer	N	Y	(Adelaide, et al., 2000; Boeva, et al., 2010)	(Rose, et al., 2010)	610084	26011

This table details all of the genes into which insertions were validated which have been previously associated with cancer. The type of cancer as well as the pertinent references for each example are listed in the table. In addition to being associated with cancer, variation in many of the genes has also been related to smoking and the relevant references are listed.

**Supp. References**

- Adelaide J, Chaffanet M, Mozziconacci MJ, Popovici C, Conte N, Fernandez F, Sobol H, Jacquemier J, Pebusque M, Ron D and others. 2000. Translocation and coamplification of loci from chromosome arms 8p and 11q in the MDA-MB-175 mammary carcinoma cell line. *Int J Oncol* 16(4):683-8.
- Bignone PA, Lee KY, Liu Y, Emilion G, Finch J, Soosay AER, Charnock FML, Beck S, Dunham I, Mungall AJ and others. 2007. RPS6KA2, a putative tumour suppressor gene at 6q27 in sporadic epithelial ovarian cancer. *Oncogene* 26(5):683-700.
- Boeva V, Jouannet S, Daveau R, Combaret V, Pierre-Eugene C, Cazes A, Louis-Brennetot C, Schleiermacher G, Ferrand S, Pierron G and others. 2013. Breakpoint features of genomic rearrangements in neuroblastoma with unbalanced translocations and chromothripsis. *Plos One* 8(8):e72182.
- Bonne A, Vreede L, Kuiper RP, Bodmer D, Jansen C, Eleveld M, van Erp F, Arkesteijn G, Hoogerbrugge N, van Ravenswaaij C and others. 2007. Mapping of constitutional translocation breakpoints in renal cell cancer patients: identification of KCNIP4 as a candidate gene. *Cancer Genet Cytogenet* 179(1):11-8.
- Burgoyne AM, Phillips-Mason PJ, Burden-Gulley SM, Robinson S, Sloan AE, Miller RH, Brady-Kalnay SM. 2009. Proteolytic Cleavage of Protein Tyrosine Phosphatase mu Regulates Glioblastoma Cell Migration. *Cancer Res* 69(17):6960-6968.
- Cao YN, He MH, Gao ZB, Peng Y, Li YL, Li L, Zhou WW, Li XC, Zhong X, Lei YM and others. 2014. Activating hotspot L205R mutation in PRKACA and adrenal Cushing's syndrome. *Science* 344(6186):913-917.

- Cha JD, Kim HJ, Cha IH. 2011. Genetic alterations in oral squamous cell carcinoma progression detected by combining array-based comparative genomic hybridization and multiplex ligation-dependent probe amplification. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* 111(5):594-607.
- Cheung KF, Lam CNY, Wu KC, Ng EKO, Chong WWS, Cheng ASL, To KF, Fan DM, Sung JJY, Yu J. 2012. Characterization of the gene structure, functional significance, and clinical application of RNF180, a novel gene in gastric cancer. *Cancer* 118(4):947-959.
- Davidson B, Abeler VM, Forsund M, Holth A, Yang Y, Kobayashi Y, Chen L, Kristensen GB, Shih Ie M, Wang TL. 2014. Gene expression signatures of primary and metastatic uterine leiomyosarcoma. *Hum Pathol* 45(4):691-700.
- Deng J, Liang H, Ying G, Zhang R, Wang B, Yu J, Fan D, Hao X. 2014. Methylation of CpG sites in RNF180 DNA promoter prediction poor survival of gastric cancer. *Oncotarget* 5(10):3173-83.
- Ding L, Getz G, Wheeler DA, Mardis ER, McLellan MD, Cibulskis K, Sougnez C, Greulich H, Muzny DM, Morgan MB and others. 2008. Somatic mutations affect key pathways in lung adenocarcinoma. *Nature* 455(7216):1069-1075.
- Dong J, Hu ZB, Wu C, Guo H, Zhou BS, Lv JC, Lu DR, Chen KX, Shi YY, Chu MJ and others. 2012. Association analyses identify multiple new lung cancer susceptibility loci and their interactions with smoking in the Chinese population. *Nature Genetics* 44(8):895-+.

Dulak AM, Stojanov P, Peng SY, Lawrence MS, Fox C, Stewart C, Bandla S, Imamura Y, Schumacher SE, Shefler E and others.

2013. Exome and whole-genome sequencing of esophageal adenocarcinoma identifies recurrent driver events and mutational complexity. *Nature Genetics* 45(5):478-U37.

Dymicka-Piekarska V, Kemono H. 2009. Does colorectal cancer clinical advancement affect adhesion molecules (sP- selectin, sE-

selectin and ICAM-1) concentration? *Thrombosis Research* 124(1):80-83.

Eckel-Passow JE, Serie DJ, Bot BM, Joseph RW, Chevillie JC, Parker AS. 2014. ANKS1B is a smoking-related molecular alteration in

clear cell renal cell carcinoma. *Bmc Urology* 14.

Fernandez-Rozadilla C, Cazier JB, Tomlinson IP, Carvajal-Carmona LG, Palles C, Lamas MJ, Baiget M, Lopez-Fernandez LA, Brea-

Fernandez A, Abuli A and others. 2013. A colorectal cancer genome-wide association study in a Spanish cohort identifies two variants associated with colorectal cancer risk at 1p33 and 8p12. *BMC Genomics* 14.

Figueiredo JC, Hsu L, Hutter CM, Lin Y, Campbell PT, Baron JA, Berndt SI, Jiao S, Casey G, Fortini B and others. 2014. Genome-

wide diet-gene interaction analyses for risk of colorectal cancer. *PLoS Genet* 10(4):e1004228.

Fleming JL, Dworkin AM, Allain DC, Fernandez S, Wei L, Peters SB, Iwenofu OH, Ridd K, Bastian BC, Toland AE. 2014. Allele-

specific imbalance mapping identifies HDAC9 as a candidate gene for cutaneous squamous cell carcinoma. *International Journal of Cancer* 134(1):244-248.



- Gonzalez-Arriagada WA, Ramos LMA, Silva AA, Vargas PA, Della Coletta R, Bingle L, Lopes MA. 2015. Salivary BPIFA1 (SPLUNC1) and BPIFA2 (SPLUNC2 A) are modified by head and neck cancer radiotherapy. *Oral Surgery Oral Medicine Oral Pathology Oral Radiology* 119(1):48-58.
- Jones S, Zhang XS, Parsons DW, Lin JCH, Leary RJ, Angenendt P, Mankoo P, Carter H, Kamiyama H, Jimeno A and others. 2008. Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science* 321(5897):1801-1806.
- Kan ZY, Jaiswal BS, Stinson J, Janakiraman V, Bhatt D, Stern HM, Yue P, Haverty PM, Bourgon R, Zheng JB and others. 2010. Diverse somatic mutation patterns and pathway alterations in human cancers. *Nature* 466(7308):869-U103.
- Kerns SL, Stone NN, Stock RG, Rath L, Ostrer H, Rosenstein BS. 2013. A 2-Stage Genome-Wide Association Study to Identify Single Nucleotide Polymorphisms Associated with Development of Urinary Symptoms After Radiotherapy for Prostate Cancer. *Journal of Urology* 190(1):102-108.
- Kitamura K, Seike M, Okano T, Matsuda K, Miyanaga A, Mizutani H, Noro R, Minegishi Y, Kubota K, Gemma A. 2014. MiR-134/487b/655 cluster regulates TGF-beta-induced epithelial-mesenchymal transition and drug resistance to gefitinib by targeting MAGI2 in lung adenocarcinoma cells. *Molecular Cancer Therapeutics* 13(2):444-53.
- Koboldt DC, Fulton RS, McLellan MD, Schmidt H, Kalicki-Veizer J, McMichael JF, Fulton LL, Dooling DJ, Ding L, Mardis ER and others. 2012. Comprehensive molecular portraits of human breast tumours. *Nature* 490(7418):61-70.

- Kohler S, Ullrich S, Richter U, Schumacher U. 2010. E-/P-selectins and colon carcinoma metastasis: first in vivo evidence for their crucial role in a clinically relevant model of spontaneous metastasis formation in the lung. *British Journal of Cancer* 102(3):602-609.
- Lauc G HJ, Pučić M, Zgaga L, Adamczyk B, Mužinić A, Novokmet M, Polašek O, Gornik O, Krištić J, Keser T, Vitart V, Scheijen B, Uh HW, Molokhia M, Patrick AL, McKeigue P, Kolčić I, Lukić IK, Swann O, van Leeuwen FN, Ruhaak LR, Houwing-Duistermaat JJ, Slagboom PE, Beekman M, de Craen AJ, Deelder AM, Zeng Q, Wang W, Hastie ND, Gyllensten U, Wilson JF, Wuhrer M, Wright AF, Rudd PM, Hayward C, Aulchenko Y, Campbell H, Rudan I. 2013. Loci Associated with N-Glycosylation of Human Immunoglobulin G Show Pleiotropy with Autoimmune Diseases and Haematological Cancers. *PLOS Genetics* 9(1).
- Li H, Yu B, Li J, Su L, Yan M, Zhu Z, Liu B. 2014. Overexpression of lncRNA H19 enhances carcinogenesis and metastasis of gastric cancer. *Oncotarget* 5(8):2318-29.
- Matta A, Tripathi SC, DeSouza LV, Grigull J, Kaur J, Chauhan SS, Srivastava A, Thakar A, Shukla NK, Duggal R and others. 2009. Heterogeneous ribonucleoprotein K is a marker of oral leukoplakia and correlates with poor prognosis of squamous cell carcinoma. *International Journal of Cancer* 125(6):1398-1406.
- Milosevic N, Kuhnemuth B, Muhlberg L, Ripka S, Griesmann H, Lolkes C, Buchholz M, Aust D, Pilarsky C, Krug S and others. 2013. Synthetic Lethality Screen Identifies RPS6KA2 as Modifier of Epidermal Growth Factor Receptor Activity in Pancreatic Cancer. *Neoplasia* 15(12):1354-+.

- Mukhopadhyay D, Jung J, Murmu N, Houchen CW, Dieckgraefe BK, Anant S. 2003. CUGBP2 plays a critical role in apoptosis of breast cancer cells in response to genotoxic injury. *Apoptosis: From Signaling Pathways to Therapeutic Tools* 1010:504-509.
- Natarajan G, Ramalingam S, Ramachandran I, May R, Queimado L, Houchen CW, Anant S. 2008. CUGBP2 downregulation by prostaglandin E2 protects colon cancer cells from radiation-induced mitotic catastrophe. *Am J Physiol Gastrointest Liver Physiol* 294(5):G1235-44.
- Rose JE, Behm FM, Drgon T, Johnson C, Uhl GR. 2010. Personalized Smoking Cessation: Interactions between Nicotine Dose, Dependence and Quit-Success Genotype Score. *Molecular Medicine* 16(7-8):247-253.
- Serra V, Eichhorn PJ, Garcia-Garcia C, Ibrahim YH, Prudkin L, Sanchez G, Rodriguez O, Anton P, Parra JL, Marlow S and others. 2013. RSK3/4 mediate resistance to PI3K pathway inhibitors in breast cancer. *J Clin Invest* 123(6):2551-63.
- Slattery ML, Lundgreen A, Herrick JS, Wolff RK. 2011. Genetic variation in RPS6KA1, RPS6KA2, RPS6KB1, RPS6KB2, and PDK1 and risk of colon or rectal cancer. *Mutation Research-Fundamental and Molecular Mechanisms of Mutagenesis* 706(1-2):13-20.
- Stransky N, Egloff AM, Tward AD, Kostic AD, Cibulskis K, Sivachenko A, Kryukov GV, Lawrence MS, Sougnez C, McKenna A and others. 2011. The Mutational Landscape of Head and Neck Squamous Cell Carcinoma. *Science* 333(6046):1157-1160.
- Sun PH, Ye L, Mason MD, Jiang WG. 2012. Protein Tyrosine Phosphatase mu (PTP mu or PTPRM), a Negative Regulator of Proliferation and Invasion of Breast Cancer Cells, Is Associated with Disease Prognosis. *Plos One* 7(11).

Tan J, Lee PL, Li Z, Jiang X, Lim YC, Hooi SC, Yu Q. 2010. B55beta-associated PP2A complex controls PDK1-directed myc signaling and modulates rapamycin sensitivity in colorectal cancer. *Cancer Cell* 18(5):459-71.

Tang H, Wei P, Duell EJ, Risch HA, Olson SH, Bueno-de-Mesquita HB, Gallinger S, Holly EA, Petersen G, Bracci PM and others. 2014. Axonal guidance signaling pathway interacting with smoking in modifying the risk of pancreatic cancer: a gene- and pathway-based interaction analysis of GWAS data. *Carcinogenesis* 35(5):1039-45.

Timson G, Banavali S, Gutierrez MI, Magrath L, Bhatia KG, Goyns MH. 2006. High level expression of N-acetylglucosamine-6-O-sulfotransferase is characteristic of a subgroup of paediatric precursor-B acute lymphoblastic leukaemia. *Cancer Letters* 242(2):239-244.

Tsaprouni LG, Yang TP, Bell J, Dick KJ, Kanoni S, Nisbet J, Vinuela A, Grundberg E, Nelson CP, Meduri E and others. 2014. Cigarette smoking reduces DNA methylation levels at multiple genomic loci but the effect is partially reversible upon cessation. *Epigenetics* 9(10):1382-1396.

Uhl GR, Drgon T, Johnson C, Walther D, David SP, Aveyard P, Murphy M, Johnstone EC, Munafò MR. 2010. Genome-wide association for smoking cessation success: participants in the Patch in Practice trial of nicotine replacement. *Pharmacogenomics* 11(3):357-67.

Uhl GR, Liu QR, Drgon T, Johnson C, Walther D, David SP, Niaura R, Lerman C, Rose E. 2008. Molecular genetics of successful smoking cessation - Convergent genome-wide association study results. *Archives of General Psychiatry* 65(6):683-693.

- Uhl GR, Liu QR, Drgon T, Johnson C, Walther D, Rose JE. 2007. Molecular genetics of nicotine dependence and abstinence: whole genome association using 520,000 SNPs. *BMC Genet* 8:10.
- Vargas PA, Speight PM, Bingle CD, Barrett AW, Bingle L. 2008. Expression of PLUNC family members in benign and malignant salivary gland tumours. *Oral Diseases* 14(7):613-619.
- Vater I, Montesinos-Rongen M, Schlesner M, Haake A, Purschke F, Sprute R, Mettenmeyer N, Nazzal I, Nagel I, Gutwein J and others. 2015. The mutational pattern of primary lymphoma of the central nervous system determined by whole-exome sequencing. *Leukemia* 29(3):677-85.
- Vazquez A, Kulkarni D, Grochola LF, Bond GL, Barnard N, Toppmeyer D, Levine AJ, Hirshfield KM. 2011. A genetic variant in a PP2A regulatory subunit encoded by the PPP2R2B gene associates with altered breast cancer risk and recurrence. *International Journal of Cancer* 128(10):2335-2343.
- Wang DD, Wang LL, Zhou J, Pan JH, Qian W, Fu JF, Zhang GL, Zhu YM, Liu CS, Wang CL and others. 2014. Reduced Expression of PTPRD Correlates with Poor Prognosis in Gastric Adenocarcinoma. *Plos One* 9(11).
- Wang LH, Tsutsumi S, Kawaguchi T, Nagasaki K, Tatsuno K, Yamamoto S, Sang F, Sonoda K, Sugawara M, Saiura A and others. 2012. Whole-exome sequencing of human pancreatic cancers and characterization of genomic instability caused by MLH1 haploinsufficiency and complete deficiency. *Genome Res* 22(2):208-219.
- Wei M, Tai G, Gao Y, Li N, Huang B, Zhou Y, Hao S, Zeng X. 2004. Modified heparin inhibits P-selectin-mediated cell adhesion of human colon carcinoma cells to immobilized platelets under dynamic flow conditions. *J Biol Chem* 279(28):29202-10.

Wen D, Xu Z, Xia L, Liu X, Tu Y, Lei H, Wang W, Wang T, Song L, Ma C and others. 2014. Important role of SUMOylation of Spliceosome factors in prostate cancer cells. *J Proteome Res* 13(8):3571-82.

Won HH, Lee J, Park JO, Park YS, Lim HY, Kang WK, Kim JW, Lee SY, Park SH. 2012. Polymorphic markers associated with severe oxaliplatin-induced, chronic peripheral neuropathy in colon cancer patients. *Cancer* 118(11):2828-36.