Office de la Propriété Intellectuelle du Canada
Un organisme d'Industrie Canada

Canadian Intellectual Property Office
An agency of Industry Canada

CA 2804335 A1 2011/12/22

(21) 2 804 335

(12) DEMANDE DE BREVET CANADIEN
CANADIAN PATENT APPLICATION

(13) A1

(86) Date de dépôt PCT/PCT Filing Date: 2010/06/15
(87) Date publication PCT/PCT Publication Date: 2011/12/22
(85) Entrée phase nationale/National Entry: 2012/12/14
(86) N° demande PCT/PCT Application No.: IB 2010/052684
(87) N° publication PCT/PCT Publication No.: 2011/158061

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

(71) Demandeurs/Applicants:
BARDELLI, ALBERTO, IT;
MARTINI, MIRIAM, IT

(72) Inventeurs/Inventors:
BARDELLI, ALBERTO, IT;
MARTINI, MIRIAM, IT

(74) Agent: SMART & BIGGAR

(54) Titre : LE GENE MLK4, UN NOUVEAU MARQUEUR DE DIAGNOSTIC ET DE PRONOSTIC DES CANCERS
(54) Title: MLK4 GENE, A NEW DIAGNOSTIC AND PROGNOSTIC MARKER IN CANCERS

(57) Abrégé/Abstract:
An in vitro diagnostic method for determining invasive potential of cancer comprising measuring MLK4 gene expression in a cancer cell sample, wherein MLK4 gene overexpression is indicative of an invasive cancer, preferably a colorectal, bladder, breast, gastric, melanoma, lung, ovary or GMB cancer.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
22 December 2011 (22.12.2011)

PCT

(10) International Publication Number
WO 2011/158061 A1

(51) International Patent Classification:
C12Q 1/68 (2006.01)

(21) International Application Number:
PCT/IB2010/052684

(22) International Filing Date:
15 June 2010 (15.06.2010)

(25) Filing Language: English

(26) Publication Language: English

(72) Inventors; and
(71) Applicants : **BARDELLI, Alberto** [IT/IT]; Via Guinicelli, 11/4, I-10132 Torino (IT). **MARTINI, Miriam** [IT/IT]; Via Sciesa, 22, I-10078 Venaria (Torino) (IT).

(74) Agent: **FREYRIA FAVA, Cristina**; c/o Buzzi, Notaro & Antonielli d'Oulx S.r.l., Via Maria Vittoria, 18, I-10123 Torino (IT).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

— *with sequence listing part of description (Rule 5.2(a))*

WO 2011/158061 A1

(54) **Title**: MLK4 GENE, A NEW DIAGNOSTIC AND PROGNOSTIC MARKER IN CANCERS

(57) **Abstract**: An in vitro diagnostic method for determining invasive potential of cancer comprising measuring MLK4 gene expression in a cancer cell sample, wherein MLK4 gene overexpression is indicative of an invasive cancer, preferably a colorectal, bladder, breast, gastric, melanoma, lung, ovary or GMB cancer.

MLK4 gene, a new diagnostic and prognostic marker in cancers

****

Technical field of the invention

The present disclosure concerns the identification of a new diagnostic and prognostic marker in cancers. More specifically, such a new marker belongs to the family of Mixed Lineage Kinases (MLKs). This marker plays a central role in the diagnosis, prognosis and the response to therapy of tumors bearing mutations in the KRAS and BRAF genes. Such tumors are characterized by an extremely poor prognosis and lack of response to therapy.

Background art

Targeting of deregulated protein kinases has proven effective for multiple cancers types. Using high-throughput mutational profiling of kinase genes, we previously identified somatic mutations in Mixed Lineage Kinase 4 (MLK4) in colorectal tumors (CRC) (1). MLK4 belongs to the family of MLK serine-threonine kinases thought to activate multiple intracellular signaling pathways. They are characterized by an amino-terminal SRC-homology domain (SH3), followed sequentially by a kinase domain, a leucine-zipper region and a Cdc42/Rac Interactive Binding (CRIB) motif. The carboxyl terminus of all MLKs is proline-rich but diverge significantly among different members of the family, indicating that these regions might serve different regulatory functions. The discovery of MLK4 somatic mutations in CRCs suggests that this kinase may be relevant for tumour initiation and development. Nothing is presently known about the biochemical and cellular properties of MLK4 in normal and neoplastic cells.

Summary of the invention

Object of the present invention is the provision of a new diagnostic and prognostic marker in cancers.

According to the present invention said object is achieved thanks to the solution having the characteristics referred to specifically in the ensuing claims. Thus the claims form integral part of the technical teaching herein provided in

relation to the present invention.

To achieve this object, the present inventors have developed an *in vitro* diagnostic method which by measuring of MLK4 gene expression, more specifically MLK4 gene overexpression in a cancer cell sample material is able to determine the potential of invasiveness of the cancer.

In an embodiment, the present disclosure concerns measuring MLK4 protein expression and/or MLK4 coding nucleic acid, preferably RNA, expression.

In a further embodiment, the present disclosure concerns measuring MLK4 gene overexpression in a bioptic sample, wherein the bioptic sample corresponds to the invasive portion of the cancer cell sample material.

In a further embodiment, the present disclosure concerns measuring wild-type MLK4 and/or mutated MLK4 gene expression, wherein the mutated MLK4 gene contains i) at least one somatic mutation at the amino acid level coding for at least one amino acid mutation selected from H261Y, H261Q, G291E, A293E, W296Stp, R470C, R553Stp, R555Stp, N596I, K629E, P843S, H890P, and M894T mutations or ii) at least one somatic mutation at the nucleic acid level selected from C781T, C783G, G872A, C878A, G888A, C1408T, C1657T, C1663T, A1787T, A1885G, C2788T, A2669C, T2681C mutation.

In a still further embodiment, the present disclosure concerns an *in vitro* diagnostic method, wherein the MLK4 gene expression measurement is performed using a reagent able to selectively bind to MLK4 protein and/or MLK4 coding nucleic acid, which is directly or indirectly labeled with a detectable substance.

In a further embodiment, the present disclosure concerns an *in vitro* diagnostic method, wherein the MLK4 gene expression measurement is performed in combination with the KRAS gene mutations and expression measurement.

In a further embodiment, the present disclosure concerns inactivation of the MLK4 protein expression or its biochemical function as a mean to therapeutically treat tumors bearings KRAS gene mutations.

Brief description of the drawings

The present invention will now be described in detail in relation to some

preferred embodiments by way of non limiting examples, referring to the annexed figures, in which:

- **Figure 1. Analysis of MLK4 in human tumors**. A) Somatic mutations identified in the MLK4 gene in CRCs and glioblastomas. B) Moderately-differentiated infiltrating colorectal carcinoma. Several neoplastic glandular structures that vary in size and shape can be seen. At the lateral border (identified by arrows) the tumor cells show intense membrane and sub-membrane staining showing infiltration. The arrowheads indicate areas of strong MLK4 expression in neoplastic cells that are infiltrating the stroma. Note that noninfiltrating cells are negative or show weak apical staining. C) High magnification of the section presented in B. D) Metastasis of CRC in a regional lymph node. The fat tissue surrounding the lymph node is infiltrated by the metastatic cells. E) Higher magnification of the section presented in C. F) Liver metastasis of colorectal carcinoma. Only the cells at the periphery of the metastatic nodule (right side) are immunostained.

- **Figure 2. Functional interaction between oncogenic RAS and MLK4 on transformation, tumor growth and response to therapy.** A) Relative kinase activity of MLK4 wt and mutant alleles. The relative kinase activity was calculated as the ratio between total MLK4 protein and the amount of ATP consumption. Results were normalized using the relative activity of wt MLK4. B) Focus forming assay in NIH3T3 cells transfected with wt and mutant MLK4 alleles. Ras (V12) is used as a positive control. **p value $\leq$ 0.01; error bars represent s.d. C) Focus forming assay in NIH3T3 cells co-transfected with mutant Ras (V12) and wt/mutant MLK4 alleles. **p value $\leq$ 0.01; error bars represent s.d. D) Biochemical validation of HKE3 cancer cells expressing the indicated MLK4 variants. E) Tumor formation by HKE3 cancer cells in xenograft mouse models expressing the indicated MLK4 variants. Cells were injected in the side of nude mice and tumor growth was measured at the indicated time points. Error bars represent s.e.m. F) Biochemical validation of DKO3 cancer cells expressing the indicated MLK4 variants. G) Tumor formation by DKO3 cancer cells expressing the indicated MLK4 variants in xenograft mouse models. Cells were injected in the side of nude mice and tumor growth was measured at the indicated time points. Error bars represent s.e.m.

- **Figure 3. Invasive, tumorigenic and metastatic potential of MLK4 mutated alleles.** A) A549 cancer cells lines were transduced with lentiviral

vectors expressing wild type or two MLK4 mutants. MLK4 expression levels were assessed by western blotting on lysates of cell transduced with empty vector or the indicated constructs. The invasive potential was assessed by migration through a matrigel coated membrane in absence or upon stimulation with Hepatocyte Growth Factor (HGF). *p value ≤ 0.05; **p value ≤ 0.01; error bars represent s.d. B) Tumor formation by A549 cancer cells expressing wild type or two MLK4 mutants in xenograft mouse models. Cells were injected in the side of nude mice and tumor growth was measured at the indicated time points. Error bars represent s.e.m. C) Upon sacrifice the lungs were labeled by airway perfusion with the India ink, and superficial metastases were counted under a stereoscopic microscope. **p value ≤ 0.01; error bars represent s.e.m. D) Invasive potential of transduced DLD1 tumor cell line. Experiments were performed as in A. *p value ≤ 0.05; **p value ≤ 0.01; error bars represent s.d. E) Tumor formation by DLD1 cancer cells expressing wild type or two MLK4 mutants in xenograft mouse models. Experiments were performed as in B. Error bars represent s.e.m. F) Analysis of DLD1-derived superficial lung metastases. Experiments were performed as in C. **p value ≤ 0.01; error bars represent s.e.m.

- **Figure 4. Downregulation or abrogation of MLK4 expression impairs the transforming and invasive potential of human cancer cells.** A) Two independent ShRNAs targeting the MLK4 gene were used to downregulate MLK4 expression in multiple cell lines (A549, Colo205, DLD1, HCT116) from different tumor types. Expression levels of the MLK4 protein were assessed by western blotting. Anchorage-independent growth (soft agar) assay performed on parental and MLK4 knock down/knock out cells. *p value ≤ 0.05; **p value ≤ 0.01; error bars represent s.d. B) Schematic representation of the vector used to KO the MLK4 genes from the genome of the HCT116 cell line. A sequential targeting strategy was used to knock out (KO) both of the MLK4 alleles. L-ITR: Left-Inverted Terminal Repeats, P: neomycin promoter, Neo: neomycin resistance cassette, pA: polyadenilation sequence, R-ITR: Right-Inverted Terminal Repeats. C) Expression levels of the MLK4 protein assessed by western blotting in parental and knock out HCT116 cells. D) The invasive potential of parental and MLK4 knock-down or KO cells was assessed by migration through a matrigel coated membrane in absence or upon stimulation with Hepatocyte Growth Factor (HGF). *p value ≤ 0.05; **p value ≤ 0.01; error bars represent s.d. E) Control and MLK4 knock-down DLD1 cancer cells were injected in the side of nude mice and tumor

growth was measured at the indicated time points. Error bars represent s.e.m. F) Control and MLK4 knock-out HCT116 cancer cells were injected in the side of nude mice and tumor growth was measured at the indicated time points. Error bars represent s.e.m.

- **Figure 5. Effects of MLK4 abrogation on the KRAS-ERK signaling pathway.** A) Measurement of 'active' (GTP bound) RAS assessed by a pull-down assay with the RAS binding domain of C-RAF in parental and MLK4 knock out cells in the absence or presence of serum. Total RAS is used as a reference for sample loading. B) Biochemical analysis of MEK1/2 and ERK1/2 kinases phosphorylation levels. Cell lysates were subjected to western blotting with the indicated antibodies. C) Biochemical analysis of p38, SAPK/JNK and AKT kinases phosphorylation levels. Cell lysates were subjected to western blotting with the indicated antibodies. D) Comparison of MEK1/2 and ERK1/2 kinases phosphorylation levels in parental, MLK4 knock out and mutant KRAS KO HCT116 cells. E) Analysis of the level of expression of SHC and Grb2 adaptor proteins detected by western-blot analysis in parental and MLK4 knock out HCT116 cells.

- **Figure 6**. A) Rates of cell proliferation of A549 and DLD1 cells transduced with wt and mutant MLK4. Error bars represent s.d. B) Cell proliferation rates of parental and MLK4 KO HCT116 cancer cells. *p value ≤ 0.05; **p value ≤ 0.01; error bars represent s.d. C) Schematic representation of the MLK4 signaling pathway according to this work.

- **Figure 7**. The colorectal cancer cell line DiFi was transduced with lentiviral vectors expressing wild type or two MLK4 mutants and RAS V12. The indicated cell lines were treated with 0.1 μg/ml of Cetuximab (Erbitux) for 72 hours, after which cell viability was assessed by the ATP assay. Results shown are normalized to untreated cells; ***p value ≤ 0.001; error bars represent s.d.

Detailed description of embodiments of the invention

The present invention will now be described in detail in relation to some preferred embodiments by way of non limiting examples.

In the following description, numerous specific details are given to provide a thorough understanding of embodiments. The embodiments can be practiced without one or more of the specific details, or with other methods, components,

materials, etc. In other instances, well-known structures, materials, or operations are not shown or described in detail to avoid obscuring aspects of the embodiments.

The headings provided herein are for convenience only and do not interpret the scope or meaning of the embodiments.

Using molecular analysis of tumor samples combined with biochemical and functional approaches we have assessed the role of MLK4 in cancer cells. We employed two complementary strategies: (i) ectopic expression of wt and mutated MLK4 alleles (forward genetics) and (ii) targeting MLK4 expression using ShRNA-mediated down regulation or somatic gene knock out (KO) of MLK4 gene (reverse genetics) in normal and cancer cell lines.

Mutational profiling of the coding region of the MLK4 gene was performed in multiple tumor types including bladder, breast, gastric, melanoma, lung, ovary and glioblastoma (GBM) samples (Table 1).

Table 1.

Tumor type	Histotype	Number of samples analysed
Bladder	**Total**	**35**
	transitional cell carcinoma	33
	cell line	2
Breast	**Total**	**88**
	ductal carcinoma	26
	lobular carcinoma	29
	medullary carcinoma	17
	mucinous carcinoma	16
Gastric	**Total**	**12**
	adenocarcinoma of gastroesophageal junction	**12**
Glioma	**Total**	**113**
	Glioblastoma	105
	anaplastic astrocytoma	2
	anaplastic oligodendroglioma	2
	high grade glioma cell lines	4
Lung	**Total**	**86**
	Adenocarcinoma	86
Melanoma	**Total**	**22**
	Primary	1
	nodal metastasis	12
	cutaneous metastasis	7
	visceral metastasis	2
Ovary	**Total**	**48**
	serous adenocarcinoma	48

In addition to those previously reported (1) we found novel somatic

mutations affecting the MLK4 gene in GBMs (Fig. 1A). Overall, the mutation frequency in CRC and GBMs were 9/192 (5%) and 4/140 (3%), respectively. Most of the non-synonymous mutations involved evolutionarily conserved amino acids some of which were affected multiple times. In addition, a subset of the mutations occurred at amino acids previously shown to be pathogenic in other protein kinases (1). Furthermore, two nonsense mutations were identified that affected arginines R553 and R555 in the C-terminus. It is possible that latter truncating mutations influence the function of MLK4 by removing the putative C-terminal inhibitory domain. Finally, the nonsense mutation (W296 stop), located in the kinase domain, may represent a "passenger", rather than a "driver" mutation.

To assess whether MLK4 deregulation occurred by mechanisms other than somatic mutations, we developed two polyclonal antibodies and used them to measure the expression pattern of MLK4 in CRC samples. Immunohistochemical (IHC) analysis was performed on a panel of 38 CRCs. The results demonstrated that most tumors expressed low or undetectable levels of MLK4 expression. However, 9 CRC samples (24%) showed clear MLK4 positivity, which was very distinctive being limited at the invasive portion of the tumor (Fig. 1B). Areas of intensely stained neoplastic cells were typically observed infiltrating the stroma (Fig. 1C). In contrast, non-infiltrating cells were negative or showed a very weak expression signal. For one sample lymph node and liver metastasis were available in addition to the primary tumour (Fig. 1 D-F); in both, MLK4 expression was again confined to the invasive portion of the tumor. DNA sequence analysis excluded that the MLK4 gene was somatically mutated in the CRC samples that demonstrated increased expression. The MLK4 expression pattern suggests that MLK4 plays a central role in tumor invasion.

Lentiviral vectors were engineered to stably transduce *MLK4* wt cDNA and two of the mutant alleles, specifically H261Y, G291E and R470C, identified in the CRC sequencing screens (Table 1). The H261Y was selected for further studies since histidine 261 was found mutated twice (H261Y and H261Q). The G291E allele was selected as it is located in a region corresponding to the BRAF V600E oncogenic mutation found in multiple tumor (2). The R470C allele was chosen as this variant has been identified in two independent CRC samples (Table 1).

Biochemical *in vitro* assay were used to assess the catalytic activity of wt and mutated MLK4. As a recipient for these experiments we used colorectal cancer cells in which the MLK4 gene was genetically inactivated and as a result the endogenous MLK4 protein was absent (see below Figure 4B). These cells were transduced with wild type or the H261Y, G291E and R470C MLK4 variants. Mutated MLK4 proteins displayed increased kinase activity as compared to the wt counterpart (Fig. 2A).

Next we sought to establish the role of MLK4 on cellular phenotypes associated with tumorigenesis using both forward and reverse genetics strategies. We initially used the standard NIH3T3 focus forming assay to assess the transforming potential of wt and mutated *MLK4* alleles. Although a positive control (*RAS* G12V) was capable of promoting focus formation at high efficiency, no effects were seen when wt or mutated MLK4 were used (Fig. 2B). These results suggest that wt and mutant *MLK4* alleles are unable, *per se*, to support cellular transformation. On the contrary, when mutant MLK4 was co-transfected in NIH3T3 cells with the oncogenic version of *RAS* (G12V) a striking synergistic effect in focus formation was observed (Fig. 2C). Importantly, two of the mutants (G291E and R470C) but not wild type MLK4 increased the transforming potential of RAS G12V.

This behavior is reminiscent of other cancer genes (such as Pokemon) whose transforming potential becomes detectable only when assessed in cooperation with other oncogenes (3).

To further assess the role of MLK4 in KRAS mediated oncogenesis, we employed an isogenic variant (HKE3) of a tumorigenic human colorectal cancer cell line, HCT116, in which the mutant *KRAS* allele is disrupted by homologous recombination (4). Consequently, HKE3 cells are a suitable model to establish whether mutated *MLK4* alleles can rescue the tumorigenic phenotype of oncogenic KRAS. HKE3 expressing wt or the MLK4 mutants (G291E and R470C) that cooperated with RAS in the transformation assays (Fig. 2D) were subcutaneously injected into nude mice. Cells expressing the empty vector were used as controls. Two weeks after inoculation, HKE3 cells transduced with the empty vector or wt *MLK4* did not generated tumors (Fig. 2E). On the contrary mice inoculated with the G291E and R470C *MLK4* mutants developed subcutaneous tumors within 2 weeks. The observation that mutated *MLK4* alleles

rescue the tumorigenic phenotype of cancer cells, in which mutant *KRAS* was deleted, supports the concept that MLK4 plays a critical role in the same pathway.

To validate this observation in a different cellular system, we used an isogenic variant (DKO3) of another highly tumorigenic human colorectal cancer cell line (DLD-1), in which the mutant *KRAS* allele has been disrupted by homologous recombination. Similar *in vivo* experiments were performed in DKO3 cells transduced with MLK4 mutant alleles with comparable results (Fig. 2F-G).

The pivotal role of the KRAS pathway in human cancers has been further underlined by the discovery that KRAS mutated colorectal tumors do not respond to the anti EGFR monoclonal antibodies cetuximab (Erbitux) and panitumumab (Vectibix) (5, WO-A-2008/112274). These findings were rapidly translated into the clinic and detection of KRAS mutations is now routinely used to select patients eligible for cetuximab therapy. We previously reported that the ectopic expression of mutated KRAS and BRAF impairs the response to cetuximab and panitumumab of colorectal cancer cells (6, 7). We reasoned that if MLK4 acts downstream to KRAS its oncogenic activation could also impairs the effects of cetuximab.

To test this hypothesis we used the DiFi colorectal cancer cell line which is inhibited by nanomolar concentration of cetuximab. In DiFi cells the KRAS pathway is activated by amplification/overexpression of the *EGFR* gene. DiFi cells are wild type for *MLK4*, *KRAS*, *BRAF*, *PIK3CA* and *PTEN* mutations, thus representing a suitable model for our experiments. Wt or mutated MLK4 proteins were expressed in DiFi cells and mutated KRAS (G12V) was used as positive control. Cells expressing mutated *MLK4* or *KRAS* alleles were less sensitive to cetuximab as indicated by their markedly increased survival upon treatment with this monoclonal antibody (Fig. 7).

These data indicate that the presence of oncogenic MLK4 can substitute for *KRAS* mutations in conferring resistance to EGFR targeted therapies, supporting the concept that KRAS and MLK4 are involved in the same signal transduction pathway (Figure 6D).

We next sought to assess the role of MLK4 alleles in *invasive-growth*, a genetic program leading to metastasis formation. To formally test this hypothesis, we evaluated whether ectopic expression of wt or mutated MLK4 might affect the invasive properties of cancer cells (A549 and DLD1). Both wt and mutant MLK4 mildly affected cell invasion measured by the matrigel Boyden chamber assays in

standard growth medium. We then assessed the invasiveness upon stimulation with the Hepatocyte Growth Factor (HGF) which has previously been linked to the *invasive growth* properties of cancer cells. In the presence of HGF, expression of both mutated *MLK4* alleles (G291E and R470C) markedly increased invasion as compared to the non-transduced cells or those expressing wt MLK4 (FiG. 3A and 3D).To determine whether MLK4 promoted cell proliferation might have contributed to the augmented invasive potential, cell proliferation rates were analyzed. Ectopic expression of MLK4 did not affect proliferation (Fig. 6A).

To further evaluate the tumorigenic potential of *MLK4* mutations, A549 and DLD1 cells expressing wt MLK4, G291E and R470C mutants were subcutaneously injected into nude mice. Cells expressing the empty vector were used as controls. Mice inoculated with wt or mutated MLK4 cells developed subcutaneous tumors within 3 weeks. Both A549 and DLD1 cells transduced with mutant MLK4 gave rise to larger tumors compared to control cells (Fig. 3B and 3E). Metastasis formation in the lungs of mice bearing MLK4 over-expressing tumors was then quantified. We found that the number of metastasis was increased in MLK4 mutant expressing tumors versus controls (Fig. 3C and 3F) in both A549 and DLD1 models.

When considered together, the forward genetic approaches indicate that mutated *MLK4* alleles cooperate or substitute oncogenic KRAS in transformation and invasion and promote metastasis formation.

As a complementary approach, reverse genetics was used to evaluate how reduced expression or deletion of the *MLK4* gene affected the tumorigenic properties of cancer cells. We first identified ShRNA that targeted the *MLK4* sequence leading to efficient downregulation of its expression. Two independent MLK4 ShRNAs were selected based on their efficiency in reducing MLK4 expression in multiple cell lines derived from colon and lung cancers, including those utilized in the forward genetic approach (Fig. 4A, upper panel). We first investigated the effect of MLK4 down-regulation on anchorage independent growth, a key feature of the neoplastic phenotype. Anchorage-independent growth of all cancer cell lines was significantly reduced or almost completely impaired (Fig. 4A).

Experiments based on ShRNA mediated down regulation may be misleading due to non-specific gene targeting. We therefore used homologous recombination to delete exon 1 (where the MLK4 kinase domain is located) from

the HCT116 cells to generate cells with permanently abrogated MLK4 expression (Fig. 4B). A two-step genetic strategy was used to obtain first heterozygous and then homozygous cells in which both alleles of the *MLK4* locus were correctly targeted. Knocked out (KO) cells were viable yet lacked MLK4 expression confirmed by PCR and immunoblotting with the anti-MLK4 antibody (Fig. 4C). Soft agar growth of MLK4 KO cells was severely impaired (Fig. 4A) confirming and substantiating the results obtained with the Sh-mediated MLK4-silencing. Similarly, the invasive potential of knocked down and KO MLK4 cells was dramatically impaired (Fig. 4D).

Next, we determined the effect of reduced or abrogated MLK4 expression on the tumorigenic potential in colorectal cancer cells *in vivo*. To this end we took advantage of DLD1 and HCT116, two colorectal cancer cell lines which harbor a single copy of the activated *KRAS* oncogene. MLK4 knock down cells and those transduced with scrambled ShRNA were evaluated for the ability to growth in xenograft experiments. When inoculated into nude mice, DLD1 control cells rapidly formed solid tumors while MLK4 knock down cells did not (Fig. 4E). To independently confirm these results we used the HCT116 cell in which *MLK4* was genetically deleted. Xenograft experiments showed that cancer cells lacking MLK4 expression were virtually unable to form subcutaneous tumors while the corresponding isogenic wt cells rapidly grew forming large tumor masses (Fig. 4F).

To gain mechanistic insights into the role of MLK4 in RAS signaling, we took advantage of wt and *MLK4* KO HCT116 cells which carry constitutively active KRAS (G13D). It has been reported that MLKs are serine/threonine protein kinases that regulate signaling by mitogen-activated-protein kinase (MAPK) pathways. We reasoned that lack of MLK4 could affect different steps of the KRAS-MAPK signaling cascade. We initially analyzed the levels of 'active' (GTP bound) KRAS and the activation (phosphorylation) of its downstream effectors MEK, ERK, JNK, p38 and AKT.

We found that in the absence of MLK4 the amount of active (GTP-bound) RAS and the levels MEK and ERK phosphorylation were reduced (Fig. 5A and 5B). The absence of MLK4 did not affect the phosphorylation levels of JNK, p38 and AKT (Fig. 5C). Of note, in HCT116 cells the knock out of MLK4 or the deletion of mutant KRAS resulted in comparable reduction of MEK and ERK phosphorylation Fig. 5D). Interesting the reduction of KRAS-MAPK activation in

MLK4 knock out cells was accompanied by lower expression of SHC to a less extent Grb2, two known amplifier of the RAS signaling cascade (Fig. 5E).

Together with the functional experiments showing KRAS-MLK4 cooperation in promoting tumorigenesis, these results indicate that the main signaling axis (the ERK/MAPK cascade) activated by oncogenic KRAS is impaired in the absence of the MLK4 kinase (Figure 5F).

A key biological output resulting from MAPK activation is the induction of cell proliferation. We therefore evaluated the growing rate of parental and KO *MLK4* cells and found that the absence of MLK4 affected the proliferation of HCT116 cells (Fig. 6B). These findings further support the concept that MLK4 is involved in the proliferation of cells that are dependent on mutant *KRAS*.

This work was prompted by the finding that *MLK4* is somatically mutated in CRCs and GBMs. Mutations of the related MLK3 kinase were recently detected in 21% of microsatellite unstable (MSI) colorectal tumors (8). We sequenced the MLK4 exons where we previously detected somatic mutations in 30 MSI colorectal tumors and did not find any mutations.

Mutant alleles of MLK4 and KRAS work in concert to trigger cellular transformation and mutant MLK4 rescues the tumorigenic potential of cells in which mutant KRAS is deleted. When considered together these findings implicate MLK4 in KRAS mediated oncogenesis. Remarkably, knock down or knock out of MLK4 abrogates the growth of cancer cells carrying mutated KRAS.

We and others have previously targeted the *MET*, *KRAS* and *PIK3CA* oncogenes by homologous recombination in HCT116 cancer cells. While targeting of each of these oncogenes drives distinct biological phenotypes, only the deletion of mutated *KRAS* and MLK4 results in abrogation of tumorigenesis.

In conclusion, these data identify the MLK4 kinase as a novel effector of KRAS, a presently undruggable human oncogene (Fig. 6C). Pharmacological targeting of the MLK4 kinase is amenable and provides a novel therapeutic prospect for KRAS mutated tumors.

Materials and Methods

Mutational Analysis

PCR primers were designed using Primer 3 (http://frodo.wi.mit.edu/cgibin/primer3/primer3_www.cgi), and synthesized by Invitrogen/Life Technologies, Inc. (Paisley, England). The PCR and sequencing primers are listed in Table 2.

Table 2.

Exon		Forward primer	Reverse primer	Sequence primer
1	1_1	CACCGATCGCGATTCCTAC SEQ ID No.:1	TCGAAGGCGACGTGTACC SEQ ID No.:2	TCCTACCCCCTCGCCTTC SEQ ID No.:3
	1_2	CTCAGCGTCCTCGTCGTC SEQ ID No.:4	CGCAGCTCGATGATGTTG SEQ ID No.:5	GCAGCTCGATGATGTTGG SEQ ID No.:6
	1_3	CATCTTCCCCGCCAACTAC SEQ ID No.:7	GGCCCAGTTGACCAGCAC SEQ ID No.:8	TCTTCCCCGCCAACTACG SEQ ID No.:9
	1_4	CAGGAGGTGGCCGTGAAG SEQ ID No.:10	CGCCCTAAGGCTGACTCC SEQ ID No.:11	CTCCCACCTCTTTCCATCTG SEQ ID No.:12
	1_5	GCACGTGCTGGTCAACTG SEQ ID No.:13	CGCCCTAAGGCTGACTCC SEQ ID No.:14	ACTGGGCCGTGCAGATAG SEQ ID No.:15
2		CAGGCACAGATAACCCACTAAAG SEQ ID No.:16	AAACCATAAAAATTCATTCAAAAAGG SEQ ID No.:17	CATGCCACAGAAAAATTGTG SEQ ID No.:18
5		GGGAATTTCAGCTTCTCCTTG SEQ ID No.:19	CCCTATCTCTTTCCTGGTTCTG SEQ ID No.:20	GGGGTCGCCAGTGTAGTG SEQ ID No.:21
6	6_1	CGATCCTGCAATATTCTTTGC SEQ ID No.:22	GTGGCACCTACGAAATGTCC SEQ ID No.:23	TTGCTGTGTGCTAAATACATGG SEQ ID No.:24
7	7_2	TTTCTCATTGTTTTCATGTTTTGAG SEQ ID No.:25	GCATGATAAATTCATCCTAGAAAATTG SEQ ID No.:26	TTTTGAGTAAGGTAGTTTCACAACTG SEQ ID No.:27
8		TTGTCAGCCCTTCAAAACTG SEQ ID No.:28	CTTGCTAATCTCCCCTCTGC SEQ ID No.:29	TGGTCTGTATCCACCAAAACC SEQ ID No.:30
9	9_1	GTAAAACGACGGCCAGT GAAAAGGGCTGCATGTGTTT SEQ ID No.:31	CTCTCTGAGCATCTTTCCCAA SEQ ID No.:32	GTAAAACGACGGCCAGT SEQ ID No.:33
	9_2	GCCTACATTGATCTACCTCTTGG SEQ ID No.:34	GTAAAACGACGGCCAGT TCTTCTCTTCCTTGGGCAAC SEQ ID No.:35	GTAAAACGACGGCCAGT SEQ ID No.:36
	9_3	CATAAAGCACAGGCTGCTGA SEQ ID No.:37	GTAAAACGACGGCCAGT AGCAGGCACTTTGTGGAGAA SEQ ID No.:38	GTAAAACGACGGCCAGT SEQ ID No.:39
	9_4	GTAAAACGACGGCCAGT CCACACCTTCTTTCTCCACAA SEQ ID No.:40	CATTCCCACATGTCTGCTGT SEQ ID No.:41	GTAAAACGACGGCCAGT SEQ ID No.:42
	9_5	GTAAAACGACGGCCAGT CAAGAAACTTGCCGTCTTCC SEQ ID No.:43	CAAATTATTAAATGTCATCACCAGGA SEQ ID No.:44	GTAAAACGACGGCCAGT SEQ ID No.:45

PCR primers that amplify the selected exons and the flanking intronic sequences, including splicing donor and acceptor regions, were used and PCR products were on average 381 bps in length. PCR was performed in both 384- and 96-well formats in 5- or 10-uL reaction volumes, respectively, containing 0.25 mmol/L deoxynucleotide triphosphates, 1 umol/L each of the forward and reverse primers, 6% DMSO, 1× PCR buffer, 1 ng/uL DNA, and 0.05 unit/uL Platinum Taq (Invitrogen/Life Technologies). A touchdown PCR program was used for PCR amplification (Peltier Thermocycler, PTC-200, MJ Research, Bio-Rad Laboratories, Inc., Italy). PCR conditions were as follows: 94°C for 2 min; three

cycles of 94°C for 15 s, 64°C for 30 s, 70°C for 30 s; three cycles of 94°C for 15 s, 61°C for 30 s, 70°C for 30 s; three cycles of 94°C for 15 s, 58°C for 30 s, 70°C for 30 s; and 35 cycles of 94°C for 15 s, 57°C for 30 s, and 70°C for 30 s, followed by 70°C for 5 min and 12°C thereafter. PCR products were purified using AMPure (Agencourt Bioscience Corp., Beckman Coulter S.p.A, Milan, Italy). Cycle sequencing was carried out using BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems, Foster City, CA) with an initial denaturation at 97°C for 3 min, followed by 28 cycles of 97°C for 10 s, 50°C for 20 s, and 60°C for 2 min.

Sequencing products were purified using CleanSeq (Agencourt Bioscience, Beckman Coulter) and analyzed on a 3730 DNA Analyzer, ABI capillary electrophoresis system (Applied Biosystems). Sequence traces were analyzed using the Mutation Surveyor software package (SoftGenetics, State College, PA).

Immunohistochemical analysis

Paraffin embedded sections were dewaxed, hydrated, treated with proteinase K (DAKO, Glostrup, Denmark) and immunostained using a labelled polymer detection system (Bond Polymer Define Detection, Vision Biosystem, Mount Waverley, Australia) and automated stainer (BOND-maX, Vision BioSystem). The primary polyclonal antibody was used at a dilution of 1:1000. Negative controls were obtained by replacement of primary antiserum with buffer. Only tumors which exhibited intense membrane were categorized as overexpressing MLK4.

DNA Constructs and Mutagenesis

The cDNA encoding for MLK4 set forth in SEQ ID NO.:46 was synthesized by RT-PCR by first inserting into the TopoTA vector. Full-length MLK4 cDNA was subcloned into the pCEV29.1 (10) or into pRRL plasmid (11). Mutants of MLK4 containing point mutations were constructed using the QuikChange II XL Site-directed mutagenesis kit (Stratagene) with MLK4 wt plasmid as the template DNA. The following oligonucleotides, and their reverse complements, were used as mutagenesis primers:

G291E,

5'-CTTTGAAGATTACAGATTTTGAGTTGGCGAGGGAATGGCAC-3’

(SEQ ID No:47);

R470C,

5'- TGAGCAGCAGCTGGCAGAGTGCGAGATCGACGTGCTGGAG-3'

(SEQ ID No.:48).

The presence of the appropriate mutations was confirmed by DNA sequencing.

Sequences of G291E and R470C mutants of MLK4 are set forth in SEQ ID No: 49 and 50, respectively.

In the following table a list of the MLK4 mutation identified in the present study are reported with the corresponding SEQ ID No.

Table 3.

MLK4 mutations Amino acid level	MLK4 mutations Nucleotide level	SEQ ID No.
H261Y	C781T	51
H261Q	C783G	52
G291E	G872A	49
A293E	C878A	53
W296Stp	G888A	54
R470C	C1408T	50
R553Stp	C1657T	55
R555Stp	C1663T	56
N596I	A1787T	57
K629E	A1885G	58
P843S	C2788T	59
H890P	A2669C	60
M894T	T2681C	61

Antibodies

MLK4 specific polyclonal antibodies were developed using a standard immunization protocol (two rabbits immunized for each epitope), we generated polyclonal antisera and antibodies directed against either the N- or the C-terminus of MLK4. Both antibodies were tested in immunoprecipitation and Western blot experiment, showing that these antibodies can be used to detect a 114 kDA protein (predicted molecular weight of MLK4) in total cell lysates. The primary antibodies used for immunoblotting were: anti-Vinculin (Sigma-Aldrich), anti-Actin (S.Cruz), anti-PERK1/2, anti-ERK1/2, anti-P-SAPK/JNK, anti-SAPK/JNK, anti-P-AKT, anti-AKT, anti-P-p38 and anti-P-38, (Cell Signaling).

Cell culture

A-549, Colo-201 and HCT116 were cultured in RPMI-1640 medium (Invitrogen), NIH3T3, DLD-1 and HCT116-HKE-3 derived clones were grown in DMEM (Invitrogen). HTERT-HME1 were cultured in growth medium containing DMEM/F-12 (Invitrogen) supplemented with 20 ng/mL EGF, 10 ug/ml insulin, and 100 ug/ml hydrocortisone, while DiFi cells were grown in F-12 (Invitrogen). Cultures of mammalian cells were maintained in the appropriate medium supplemented with 10% fetal bovine serum (Sigma-Aldrich), 2 mM Lglutamine, 100 units/ml penicillin/streptomycin at 37°C in humidified air with 5% CO_2. The mouse NIH3T3 fibroblasts are available by ATCC (CRL-1658) and were cultured in DMEM supplemented with 10% bovine serum (heat-inactivated at 56°C for 30 min) (Invitrogen) and 2 mM L-glutamine. All the cell lines are publicly available by American Type Culture Collection (ATCC), Manassas, USA.

Transfections and transformation assays

NIH3T3 fibroblasts were seeded at 2 x 10^5 cells in 100 mm plate and grown for 24 h before transfection using high-efficiency liposome transfection method (Lipofectamine 2000 and Plus Reagent; Invitrogen). Each plate was transfected with 4 μg of the pCEV29.1 vector containing either no insert or a cDNA encoding for a MLK4 protein (wild-type or mutant). For co-operation experiments, cells were transiently cotransfected with 0.1 μg of RasV12 plasmid and 4μg of pCEV containing the different MLK4 wt and mutants. Two days after transfection, cells were split into three plates cultured in DMEM containing 5% Bovine Serum and used for measuring focus formation. Foci were scored two weeks after transfection following fixation with glutharaldeyde 11% and Giemsa staining.

Protein analysis

Total cellular proteins were extracted by solubilizing the cells in boiling SDS buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, and 1% SDS). The amount of proteins was quantified by the BCA Protein Assay Reagent kit (Pierce Chemical Co.) Extracts were electrophoresed on SDS-polyacrylamide gels, transferred onto nitrocellulose membranes (Hybond; GE Healthcare) and blocked in phosphate buffered saline, 0.1% Tween 20, 5% BSA. The membrane was then

incubated with appropriate dilutions of primary antibody (either MLK4, vinculin, actin, pERK, ERK, pJNK, JNK, pAKT, AKT,pP38, P38), followed by the appropriate peroxidase conjugated secondary antibody (Donkey anti Rabbit Bio-Rad Laboratories). Final detection was done by enhanced chemiluminescence (GE Healthcare). For immunoprecipitations, cells were lysed at 4°C with 800µl of a buffer containing 20 mM Tris-HCl [pH 7.4], 150 mM NaCl, 10% glycerol, and 1% Triton X-100 in the presence of protease and phosphatase inhibitors (1 mM Na_3VO_4, 100 mM NaF, 1 mM PMSF, 10 µg/ml leupeptin, 10 µg/ml aprotinin, and 1 µg/ml pepstatin). Extracts were clarified at 12,000 rpm for 20 min and quantified by bicinchoninic assay. Protein A-Sepharose beads were pre-incubated with primary antibody (1h at +4°C) and then washed three times with lysis buffer before incubation with total cellular proteins. Extracts were incubated with primary antibody for 2 h at +4°C. Immune complexes were collected with protein A–Sepharose, washed three times in lysis buffer. SDS PAGE western blotting was performed as previously described.

Ras pull-down assay

Cell lysates were incubated with GST–RBD on glutathione–agarose beads at 4°C for 1h. The beads were then collected by centrifugation, resuspended in Laemmli reducing sample buffer and boiled for 8 minutes. Western blot analysis was performed using anti-Ras antibody (Calbiochem).

Kinase assays

Immunoprecipitated MLK4 was subjected to two independent strategies. For radioactive kinase assay, immunoprecipitates were washed three times with kinase assay buffer [50 mM Hepes (pH 7.5), 150 mM NaCl, 12.5 mM $MgCl_2$, 2mM EGTA 1mM dithiothreitol and 1 mM sodium orthovanadate] and the incubated with kinase buffer in the presence of 15 µM ATP including [γ-^{32}P]ATP for 30 min at 37°C. Reactions were stopped by the addition of 30 µl of Laemmli buffer and boiling for 1 min. Samples were analyzed by 15% SDS/PAGE followed by autoradiography.

Suppression of gene expression by RNAi

MLK4 expression was suppressed in tumor cells by lentiviral-mediated expression of shRNAs specifically targeting the MLK4 transcript. To rule out the

variability of biological responses, we used four different shRNA directed against MLK4 sequence. ShRNA sequences were derived from the Sigma-Aldrich Mission library under accession no. NM_032435 (sequence 1, TRCN0000003209; sequence 2, TRCN0000003210; sequence 3, TRCN0000003211; sequence 4, TRCN0000003213). All sequences were able to suppress MLK4 levels while control ShCTRL (SHC002V) was not. The two most efficient ShRNA, among the different recipient cell lines, were then selected to be used in the subsequent experiments. Western blotting analysis of shRNA infected cells showed a reduction by over 80% of MLK4 protein. Results shown in the figures refer to sequence 2 and 4, renamed as Sh1 and Sh2 respectively.

Targeted deletion of the MLK4 locus in human cancer cells

Disruption of the MLK4 exon 1 in the CRC HCT116 cells was performed as previously described (12). The targeting vector pAAV-Neo-MLK4 was constructed by PCR; DLD1 genomic was used as template for both homology arms. Constructs and primer sequences are publicly available at Università degli Studi di Torino - Dipartimento Scienze Oncologiche, strada Provinciale 142, Candiolo (TO). Clones were selected after 2 weeks of growth under 0.4 mg/ml for HCT116 cells geneticin (Invitrogen, Carlsbad, CA) selection and then propagated in the absence of selective agents. Homologous recombination events were identified by locus-specific PCR screening.

Anchorage-independent growth assays

Cells were diluted to a concentration of 500-1000 cells/ml in appropriate medium containing 5% FBS, 0.5% Seaplaque agar. Cells were seeded in 24-well plates (1 ml/well) containing a 1 % agar underlay and supplemented three times a week with corresponding medium. Colonies were photographed and scored two weeks after seeding.

Invasion Assays

Cell migration and invasion assays were performed using Costar 24-well plate transwell either 8 μm pore size transwell migration plates or Matrigel matrix-coated polycarbonate filters, respectively. Lower wells containing 500 μl of media to which either 5% serum (to measure basal condition) or 10ng/ml HGF

(to measure induced invasion) had been added, whereas upper wells contained 100 µl of cells, suspended to a concentration of 10^5 for HCT116 or 5×10^4 for A549 and DLD1, in media containing the same serum concentration of lower wells. After incubation for 20 hr, cells/media in the upper wells were discarded, and the migrated cells on the underside of the transwell membranes were fixed with glutaraldeide 11% in PBS, stained with Crystal Violet and counted with Metamorph Image Analysis Software.

Drug viability assay

Cetuximab was obtained from the Hospital Pharmacy. Cell lines were seeded in 100µl medium at appropriate density in 96-well plastic culture plates. Plates were incubated at 37°C in 5% CO_2 for 72h, after which cell viability was assessed by ATP content using the CellTiter-Glo® Luminescent Assay (Promega). Luminescence was detected using a DTX-880 plate reader (Beckman-Coulter).

Animal studies

All animal procedures were approved by the Ethical Commission of the University of Turin and by the Italian Ministry of Health. Six-week-old immunocompromised CD1–/– nude athymic female mice (Charles River Laboratories, Lecco, Italy) were injected subcutaneously in right posterior flanks. Tumor appearance was evacuate every 2d using a caliper. Tumor volume was calculated using the formula V = 4/3 x (d/2)2 x (D/2), where d is the minor tumor axis and D is the major tumor axis. Superficial pulmonary metastases were contrasted by black India ink air way infusion before excision, and were counted on dissected lung lobes under a stereoscopic microscope.

Cell lines were seeded in 100µl medium at appropriate density in 96-well plastic culture plates. Plates were incubated with cetuximab (Erbitux) at 37°C in 5% CO_2 for 72h, after which cell viability was assessed by ATP content using the CellTiter-Glo® Luminescent Assay (Promega). Luminescence was detected using a DTX-880 plate reader (Beckman-Coulter).

SEQUENCE LISTING

<110> Bardelli, Alberto
Martini, Miriam

<120> MLK4 gene, a new diagnostic and prognostic marker in cancers

<130> BWO12029-CF

<160> 61

<170> PatentIn version 3.5

<210> 1
<211> 19
<212> DNA
<213> artificial

<220>
<223> forward primer exon 1_1

<400> 1
caccgatcgc gattcctac 19

<210> 2
<211> 18
<212> DNA
<213> artificial

<220>
<223> reverse primer exon 1_1

<400> 2
tcgaaggcga cgtgtacc 18

<210> 3
<211> 18
<212> DNA
<213> artificial

<220>
<223> sequence primer exon 1_1

<400> 3
tcctaccccc tcgccttc 18

<210> 4
<211> 18
<212> DNA
<213> artificial

<220>
<223> forward primer exon 1_2

<400> 4
ctcagcgtcc tcgtcgtc 18

<210> 5
<211> 18
<212> DNA
<213> artificial

<220>
<223> reverse primer exon 1_2

<400> 5
cgcagctcga tgatgttg 18

<210> 6
<211> 18
<212> DNA
<213> artificial

<220>
<223> sequence primer exon 1_2

<400> 6
gcagctcgat gatgttgg 18

<210> 7
<211> 19
<212> DNA
<213> artificial

<220>
<223> forward primer exon 1_3

<400> 7
catcttcccc gccaactac 19

<210> 8
<211> 18
<212> DNA
<213> artificial

<220>
<223> reverse primer exon 1_3

<400> 8
ggcccagttg accagcac 18

<210> 9
<211> 18
<212> DNA
<213> artificial

<220>
<223> sequence primer exon 1_3

<400> 9
tcttccccgc caactacg 18

<210> 10
<211> 18
<212> DNA
<213> artificial

<220>
<223> forward primer exon 1_4

<400> 10
caggaggtgg ccgtgaag 18

<210> 11
<211> 18
<212> DNA
<213> artificial

<220>
<223> reverse primer exon 1_4

<400> 11
cgccctaagg ctgactcc 18

<210> 12
<211> 20
<212> DNA
<213> artificial

<220>
<223> sequence primer exon 1_4

<400> 12
ctcccacctc tttccatctg 20

<210> 13
<211> 18
<212> DNA
<213> artificial

<220>
<223> forward primer exon 1_5

<400> 13
gcacgtgctg gtcaactg 18

<210> 14
<211> 18
<212> DNA
<213> artificial

<220>
<223> reverse primer exon 1_5

<400> 14
cgccctaagg ctgactcc 18

<210> 15
<211> 18
<212> DNA
<213> artificial

<220>
<223> sequence primer exon 1_5

<400> 15
actgggccgt gcagatag 18

<210> 16
<211> 23
<212> DNA
<213> artificial

<220>
<223> forward primer exon 2

<400> 16
caggcacaga taacccacta aag 23

<210> 17
<211> 26
<212> DNA
<213> artificial

<220>

<223> reverse primer exon 2

<400> 17
aaaccataaa aattcattca aaaagg 26

<210> 18
<211> 20
<212> DNA
<213> artificial

<220>
<223> sequence primer exon 2

<400> 18
catgccacag aaaaattgtg 20

<210> 19
<211> 21
<212> DNA
<213> artificial

<220>
<223> forward primer exon 5

<400> 19
gggaatttca gcttctcctt g 21

<210> 20
<211> 22
<212> DNA
<213> artificial

<220>
<223> reverse primer exon 5

<400> 20
ccctatctct ttcctggttc tg 22

<210> 21
<211> 18
<212> DNA
<213> artificial

<220>
<223> sequence primer exon 5

<400> 21
ggggtcgcca gtgtagtg 18

<210> 22
<211> 21
<212> DNA
<213> artificial

<220>
<223> forward primer exon 6_1

<400> 22
cgatcctgca atattctttg c 21

<210> 23
<211> 20
<212> DNA
<213> artificial

<220>
<223> reverse primer exon 6_1

<400> 23
gtggcaccta cgaaatgtcc 20

<210> 24
<211> 22
<212> DNA
<213> artificial

<220>
<223> sequence primer exon 6_1

<400> 24
ttgctgtgtg ctaaatacat gg 22

<210> 25
<211> 25
<212> DNA
<213> artificial

<220>
<223> forward primer exon 7_2

<400> 25
tttctcattg ttttcatgtt ttgag 25

<210> 26
<211> 27
<212> DNA
<213> artificial

<220>
<223> reverse primer exon 7_2

<400> 26
gcatgataaa ttcatcctag aaaattg 27

<210> 27
<211> 26
<212> DNA
<213> artificial

<220>
<223> sequence primer exon 7_2

<400> 27
ttttgagtaa ggtagtttca caactg 26

<210> 28
<211> 20
<212> DNA
<213> artificial

<220>
<223> forward primer exon 8

<400> 28
ttgtcagccc ttcaaaactg 20

<210> 29
<211> 20
<212> DNA
<213> artificial

<220>
<223> reverse primer exon 8

<400> 29
cttgctaatc tcccctctgc 20

<210> 30
<211> 21
<212> DNA
<213> artificial

<220>
<223> sequence primer exon 8

<400> 30
tggtctgtat ccaccaaaac c 21

<210> 31
<211> 37
<212> DNA
<213> artificial

<220>
<223> forward primer exon 9_1

<400> 31
gtaaaacgac ggccagtgaa aagggctgca tgtgttt 37

<210> 32
<211> 21
<212> DNA
<213> artificial

<220>
<223> reverse primer exon 9_1

<400> 32
ctctctgagc atctttccca a 21

<210> 33
<211> 17
<212> DNA
<213> artificial

<220>
<223> sequence primer exon 9_1

<400> 33
gtaaaacgac ggccagt 17

<210> 34
<211> 23
<212> DNA
<213> artificial

<220>
<223> forward primer exon 9_2

<400> 34
gcctacattg atctacctct tgg 23

<210> 35
<211> 37
<212> DNA

<213> artificial

<220>
<223> reverse primer exon 9_2

<400> 35
gtaaaacgac ggccagttct tctcttcctt gggcaac 37

<210> 36
<211> 17
<212> DNA
<213> artificial

<220>
<223> sequence primer exon 9_2

<400> 36
gtaaaacgac ggccagt 17

<210> 37
<211> 20
<212> DNA
<213> artificial

<220>
<223> forward primer exon 9_3

<400> 37
cataaagcac aggctgctga 20

<210> 38
<211> 37
<212> DNA
<213> artificial

<220>
<223> reverse primer exon 9_3

<400> 38
gtaaaacgac ggccagtagc aggcactttg tggagaa 37

<210> 39
<211> 17
<212> DNA
<213> artificial

<220>
<223> sequence primer exon 9_3

<400> 39
gtaaaacgac ggccagt 17

<210> 40
<211> 38
<212> DNA
<213> artificial

<220>
<223> forward primer exon 9_4

<400> 40
gtaaaacgac ggccagtcca caccttcttt ctccacaa 38

<210> 41
<211> 20

<212> DNA
<213> artificial

<220>
<223> reverse primer exon 9_4

<400> 41
cattcccaca tgtctgctgt 20

<210> 42
<211> 17
<212> DNA
<213> artificial

<220>
<223> sequence primer exon 9_4

<400> 42
gtaaaacgac ggccagt 17

<210> 43
<211> 37
<212> DNA
<213> artificial

<220>
<223> forward primer exon 9_5

<400> 43
gtaaaacgac ggccagtcaa gaaacttgcc gtcttcc 37

<210> 44
<211> 26
<212> DNA
<213> artificial

<220>
<223> reverse primer exon 9_5

<400> 44
caaattatta aatgtcatca ccagga 26

<210> 45
<211> 17
<212> DNA
<213> artificial

<220>
<223> sequence primer exon 9_5

<400> 45
gtaaaacgac ggccagt 17

<210> 46
<211> 3111
<212> DNA
<213> homo sapiens

<400> 46
atggctttgc ggggcgccgc gggagcgacc gacaccccgg tgtcctcggc cgggggagcc 60

cccggcggct cagcgtcctc gtcgtccacc tcctcgggcg gctcggcctc ggcgggcgcg 120

gggctgtggg ccgcgctcta tgactacgag gctcgcggcg aggacgagct gagcctgcgg 180

cgcggccagc tggtggaggt gctgtcgcag gacgccgccg tgtcgggcga cgagggctgg 240

tgggcaggcc aggtgcagcg gcgcctcggc atcttccccg ccaactacgt ggctccctgc 300

cgcccggccg ccagccccgc gccgccgccc tcgcggccca gctccccggt acacgtcgcc 360

ttcgagcggc tggagctgaa ggagctcatc ggcgctgggg gcttcgggca ggtgtaccgc 420

gccacctggc agggccagga ggtggccgtg aaggcggcgc gccaggaccc ggagcaggac 480

gcggcggcgg ctgccgagag cgtgcggcgc gaggctcggc tcttcgccat gctgcggcac 540

cccaacatca tcgagctgcg cggcgtgtgc ctgcagcagc cgcacctctg cctggtgctg 600

gagttcgccc gcggcggagc gctcaaccga gcgctggccg ctgccaacgc cgccccggac 660

ccgcgcgcgc ccggcccccg ccgcgcgcgc cgcatccctc cgcacgtgct ggtcaactgg 720

gccgtgcaga tagcgcgggg catgctctac ctgcatgagg aggccttcgt gcccatcctg 780

caccgggacc tcaagtccag caacattttg ctacttgaga agatagaaca tgatgacatc 840

tgcaataaaa ctttgaagat tacagatttt gggttggcga gggaatggca caggaccacc 900

aaaatgagca cagcaggcac ctatgcctgg atggcccccg aagtgatcaa gtcttccttg 960

ttttctaagg gaagcgacat ctggagctat ggagtgctgc tgtgggaact gctcaccgga 1020

gaagtcccct atcggggcat tgatggcctc gccgtggctt atggggtagc agtcaataaa 1080

ctcactttgc ccattccatc cacctgccct gagccgtttg ccaagctcat gaaagaatgc 1140

tggcaacaag accctcatat tcgtccatcg tttgccttaa ttctcgaaca gttgactgct 1200

attgaagggg cagtgatgac tgagatgcct caagaatctt ttcattccat gcaagatgac 1260

tggaaactag aaattcaaca aatgtttgat gagttgagaa caaaggaaaa ggagctgcga 1320

tcccgggaag aggagctgac tcgggcggct ctgcagcaga agtctcagga ggagctgcta 1380

aagcggcgtg agcagcagct ggcagagcgc gagatcgacg tgctggagcg ggaacttaac 1440

attctgatat tccagctaaa ccaggagaag cccaaggtaa agaagaggaa gggcaagttt 1500

aagagaagtc gtttaaagct caaagatgga catcgaatca gtttaccttc agatttccag 1560

cacaagataa ccgtgcaggc ctctcccaac ttggacaaac ggcggagcct gaacagcagc 1620

agttccagtc ccccgagcag ccccacaatg atgccccgac tccgagccat acagttgact 1680

tcagatgaaa gcaataaaac ttggggaagg aacacagtct ttcgacaaga agaatttgag 1740

gatgtaaaaa ggaattttaa gaaaaaaggt tgtacctggg gaccaaattc cattcaaatg 1800

aaagatagaa cagattgcaa agaaaggata agacctctct ccgatggcaa cagtccttgg 1860

tcaactatct taataaaaaa tcagaaaacc atgcccttgg cttcattgtt tgtggaccag 1920

ccagggtcct gtgaagagcc aaaactttcc cctgatggat tagaacacag aaaaccaaaa 1980

caaataaaat tgcctagtca ggcctacatt gatctacctc ttgggaaaga tgctcagaga 2040

gagaatcctg cagaagctga aagctgggag gaggcagcct ctgcgaatgc tgccacagtc 2100

tccattgaga tgactcctac gaatagtctg agtagatccc cccagagaaa gaaaacggag 2160

tcagctctgt atgggtgcac cgtccttctg gcatcggtgg ctctgggact ggacctcaga 2220

gagcttcata aagcacaggc tgctgaagaa ccgttgccca aggaagagaa gaagaaacga 2280

gagggaatct tccagcgggc ttccaagtcc cgcagaagcg ccagtcctcc cacaagcctg 2340

ccatccacct gtggggaggc cagcagccca ccctccctgc cactgtcaag tgccctgggc 2400

atcctctcca caccttcttt ctccacaaag tgcctgctgc agatggacag tgaagatcca 2460

ctggtggaca gtgcacctgt cacttgtgac tctgagatgc tcactccgga tttttgtccc 2520

actgccccag gaagtggtcg tgagccagcc ctcatgccaa gacttgacac tgattgtagt 2580

gtatcaagaa acttgccgtc ttccttccta cagcagacat gtgggaatgt accttactgt 2640

gcttcttcaa aacatagacc gtcacatcac agacggacca tgtctgatgg aaatccgacc 2700

ccaactggtg caactattat ctcagccact ggagcctctg cactgccact ctgcccctca 2760

cctgctcctc acagtcatct gccaagggag gtctcaccca agaagcacag cactgtccac 2820

atcgtgcctc agcgtcgccc tgcctccctg agaagccgct cagatctgcc tcaggcttac 2880

ccacagacag cagtgtctca gctggcacag actgcctgtg tagtgggtcg cccaggacca 2940

catcccaccc aattcctcgc tgccaaggag agaactaaat cccatgtgcc ttcattactg 3000

gatgctgacg tggaaggtca gagcagggac tacactgtgc cactgtgcag aatgaggagc 3060

aaaaccagcc ggccatctat atatgaactg gagaaagaat tcctgtctta a 3111

<210> 47
<211> 41
<212> DNA
<213> artificial

<220>
<223> mutagenesis primer G291E

<400> 47
ctttgaagat tacagatttt gagttggcga gggaatggca c 41

<210> 48
<211> 40
<212> DNA
<213> artificial

<220>
<223> mutagenesis primer R470C

<400> 48
tgagcagcag ctggcagagt gcgagatcga cgtgctggag 40

<210> 49
<211> 3041
<212> DNA
<213> homo sapiens

<400> 49
atggctttgc ggggcgccgc gggagcgacc gacaccccgg tgtcctcggc cgggggagcc 60

cccggcggct cagcgtcctc gtcgtccacc tcctcgggcg gctcggcctc ggcgggcgcg 120

gggctgtggg ccgcgctcta tgactacgag gctcgcggcg aggacgagct gagcctgcgg 180

cgcggccagc tggtggaggt gctgtcgcag gacgccgccg tgtcgggcga cgagggctgg 240

tgggcaggcc aggtgcagcg gcgcctcggc atcttccccg ccaactacgt ggctccctgc 300

cgcccggccg ccagccccgc gccgccgccc tcgcggccca gctccccggt acacgtcgcc 360

ttcgagcggc tggagctgaa ggagctcatc ggcgctgggg gcttcgggca ggtgtaccgc 420

gccacctggc agggccagga ggtggccgtg aaggcggcgc gccaggaccc ggagcaggac 480

gcggcggcgg ctgccgagag cgtgcggcgc gaggctcggc tcttcgccat gctgcggcac 540

cccaacatca tcgagctgcg cggcgtgtgc ctgcagcagc cgcacctctg cctggtgctg 600

gagttcgccc gcggcggagc gctcaaccga gcgctggccg ctgccaacgc cgccccggac 660

ccgcgcgcgc ccggcccccg ccgcgcgcgc cgcatccctc cgcacgtgct ggtcaactgg 720

gccgtgcaga tagcgcgggg catgctctac ctgcatgagg aggccttcgt gcccatcctg 780

caccgggacc tcaagtccag caacattttg ctacttgaga agatagaaca tgatgacatc 840

tgcaataaaa ctttgaagat tacagatttt gagttggcga gggaatggca caggaccacc 900

aaaatgagca cagcaggcac ctatgcctgg atggcccccg aagtgatcaa gtcttccttg 960

ttttctaagg gaagcgacat ctggagctat ggagtgctgc tgtgggaact gctcaccgga 1020

gaagtcccct atcggggcat tgatggcctc gccgtggctt atggggtagc agtcaataaa 1080

ctcactttgc ccattccatc cacctgccct gagccgtttg ccaagctcat gaaagaatgc 1140

tggcaacaag accctcatat tcgtccatcg tttgccttaa ttctcgaaca gttgactgct 1200

attgaagggg cagtgatgac tgagatgcct caagaatctt ttcattccat gcaagatgac 1260

tggaaactag aaattcaaca aatgtttgat gagttgagaa caaaggaaaa ggagctgcga 1320

tcccgggaag aggagctgac tcgggcggct ctgcagcaga agtctcagga ggagctgcta 1380

aagcggcgtg agcagcagct ggcagagcgc gagatcgacg tgctggagcg ggaacttaac 1440

attctgatat tccagctaaa ccaggagaag cccaaggtaa agaagaggaa gggcaagttt 1500

aagagaagtc gtttaaagct caaagatgga catcgaatca gtttaccttc agatttccag 1560

cacaagataa ccgtgcaggc ctctcccaac ttggacaaac ggcggagcct gaacagcagc 1620

agttccagtc ccccgagcag ccccacaatg atgccccgac tccgagccat acagttgact 1680

tcagatgaaa gcaataaaac ttggggaagg aacacagtct ttcgacaaga agaatttgag 1740

gatgtaaaaa ggaattttaa gaaaaaaggt tgtacctggg gaccaaattc cattcaaatg 1800

aaagatagaa cagattgcaa agaaaggata agacctctct ccgatggcaa cagtccttgg 1860

tcaactatct taataaaaaa tcagaaaacc atgcccttgg cttcattgtt tgtggaccag 1920

ccagggtcct gtgaagagcc aaaactttcc cctgatggat tagaacacag aaaaccaaaa 1980

caaataaaat tgcctagtca ggcctacatt gatctacctc ttgggaaaga tgctcagaga 2040

gagaatcctg cagaagctga aagctgggag gaggcagcct ctgcgaatgc tgccacagtc 2100

tccattgaga tgactcctac gaatagtctg agtagatccc cccagagaaa gaaaacggag 2160

tcagctctgt atgggtgcac cgtccttctg gcatcggtgg ctctgggact ggacctcaga 2220

gagcttcata aagcacaggc tgctgaagaa ccgttgccca aggaagagaa gaagaaacga 2280

gagggaatct tccagcgggc ttccaagtcc cgcagaagcg ccagtcctcc cacaagcctg 2340

ccatccacct gtggggaggc cagcagccca ccctccctgc cactgtcaag tgccctgggc 2400

atcctctcca caccttcttt ctccacaaag tgcctgctgc agatggacag tgaagatcca 2460

ctggtggaca gtgcacctgt cacttgtgac tctgagatgc tcactccgga tttttgtccc 2520

actgccccag gaagtggtcg tgagccagcc ctcatgccaa gacttgacac tgattgtagt 2580

gtatcaagaa acttgccgtc ttccttccta cagcagacat gtgggaatgt accttactgt 2640

gcttcttcaa aacatagacc gtcacatcac agacggacca tgtctgatgg aaatccgacc 2700

ccaactggtg caactattat ctcagccact ggagcctctg cactgccact ctgcccctca 2760

cctgctcctc acagtcatct gccaagggag gtctcaccca agaagcacag cactgtccac 2820

atcgtgcctc agcgtcgccc tgcctccctg agaagccgct cagatctgcc tcaggcttac 2880

ccacagacag cagtgtctca gctggcacag actgcctgtg tagtgggtcg cccaggacca 2940

catcccaccc aattcctcgc tgccaaggag agaactaaat cccatgtgcc ttcattactg 3000

gatgctgacg tggaaggtca gagcagggac tacactgtgc c 3041

<210> 50
<211> 3041
<212> DNA
<213> homo sapiens

<400> 50
atggctttgc ggggcgccgc gggagcgacc gacaccccgg tgtcctcggc cgggggagcc 60

cccggcggct cagcgtcctc gtcgtccacc tcctcgggcg gctcggcctc ggcgggcgcg 120

gggctgtggg ccgcgctcta tgactacgag gctcgcggcg aggacgagct gagcctgcgg 180

cgcggccagc tggtggaggt gctgtcgcag gacgccgccg tgtcgggcga cgagggctgg 240

tgggcaggcc aggtgcagcg gcgcctcggc atcttccccg ccaactacgt ggctccctgc 300

cgcccggccg ccagccccgc gccgccgccc tcgcggccca gctccccggt acacgtcgcc 360

ttcgagcggc tggagctgaa ggagctcatc ggcgctgggg gcttcgggca ggtgtaccgc 420

gccacctggc agggccagga ggtggccgtg aaggcggcgc gccaggaccc ggagcaggac 480

gcggcggcgg ctgccgagag cgtgcggcgc gaggctcggc tcttcgccat gctgcggcac 540

cccaacatca tcgagctgcg cggcgtgtgc ctgcagcagc cgcacctctg cctggtgctg 600

gagttcgccc gcggcggagc gctcaaccga gcgctggccg ctgccaacgc cgccccggac 660

ccgcgcgcgc ccggcccccg ccgcgcgcgc cgcatccctc cgcacgtgct ggtcaactgg 720

gccgtgcaga tagcgcgggg catgctctac ctgcatgagg aggccttcgt gcccatcctg 780

caccgggacc tcaagtccag caacattttg ctacttgaga agatagaaca tgatgacatc 840

tgcaataaaa ctttgaagat tacagatttt gggttggcga gggaatggca caggaccacc 900

aaaatgagca cagcaggcac ctatgcctgg atggcccccg aagtgatcaa gtcttccttg 960

ttttctaagg gaagcgacat ctggagctat ggagtgctgc tgtgggaact gctcaccgga 1020

gaagtcccct atcggggcat tgatggcctc gccgtggctt atggggtagc agtcaataaa 1080

ctcactttgc ccattccatc cacctgccct gagccgtttg ccaagctcat gaaagaatgc 1140

tggcaacaag accctcatat tcgtccatcg tttgccttaa ttctcgaaca gttgactgct 1200

attgaagggg cagtgatgac tgagatgcct caagaatctt ttcattccat gcaagatgac 1260

tggaaactag aaattcaaca aatgtttgat gagttgagaa caaaggaaaa ggagctgcga 1320

tcccgggaag aggagctgac tcgggcggct ctgcagcaga agtctcagga ggagctgcta 1380

aagcggcgtg agcagcagct ggcagagtgc gagatcgacg tgctggagcg ggaacttaac 1440

attctgatat tccagctaaa ccaggagaag cccaaggtaa agaagaggaa gggcaagttt 1500

aagagaagtc gtttaaagct caaagatgga catcgaatca gtttaccttc agatttccag 1560

cacaagataa ccgtgcaggc ctctcccaac ttggacaaac ggcggagcct gaacagcagc 1620

agttccagtc ccccgagcag ccccacaatg atgccccgac tccgagccat acagttgact 1680

tcagatgaaa gcaataaaac ttggggaagg aacacagtct ttcgacaaga agaatttgag 1740

gatgtaaaaa ggaattttaa gaaaaaaggt tgtacctggg gaccaaattc cattcaaatg 1800

aaagatagaa cagattgcaa agaaaggata agacctctct ccgatggcaa cagtccttgg 1860

tcaactatct taataaaaaa tcagaaaacc atgcccttgg cttcattgtt tgtggaccag 1920

ccagggtcct gtgaagagcc aaaactttcc cctgatggat tagaacacag aaaaccaaaa 1980

caaataaaat tgcctagtca ggcctacatt gatctacctc ttgggaaaga tgctcagaga 2040

gagaatcctg cagaagctga aagctgggag gaggcagcct ctgcgaatgc tgccacagtc 2100

tccattgaga tgactcctac gaatagtctg agtagatccc cccagagaaa gaaaacggag 2160

tcagctctgt atgggtgcac cgtccttctg gcatcggtgg ctctgggact ggacctcaga 2220

gagcttcata aagcacaggc tgctgaagaa ccgttgccca aggaagagaa gaagaaacga 2280

gagggaatct tccagcgggc ttccaagtcc cgcagaagcg ccagtcctcc cacaagcctg 2340

ccatccacct gtggggaggc cagcagccca ccctccctgc cactgtcaag tgccctgggc 2400

atcctctcca caccttcttt ctccacaaag tgcctgctgc agatggacag tgaagatcca 2460

ctggtggaca gtgcacctgt cacttgtgac tctgagatgc tcactccgga tttttgtccc 2520

actgccccag gaagtggtcg tgagccagcc ctcatgccaa gacttgacac tgattgtagt 2580

gtatcaagaa acttgccgtc ttccttccta cagcagacat gtgggaatgt accttactgt 2640

gcttcttcaa aacatagacc gtcacatcac agacggacca tgtctgatgg aaatccgacc 2700

ccaactggtg caactattat ctcagccact ggagcctctg cactgccact ctgcccctca 2760

cctgctcctc acagtcatct gccaagggag gtctcaccca agaagcacag cactgtccac 2820

atcgtgcctc agcgtcgccc tgcctccctg agaagccgct cagatctgcc tcaggcttac 2880

ccacagacag cagtgtctca gctggcacag actgcctgtg tagtgggtcg cccaggacca 2940

catcccaccc aattcctcgc tgccaaggag agaactaaat cccatgtgcc ttcattactg 3000

gatgctgacg tggaaggtca gagcagggac tacactgtgc c 3041

<210> 51
<211> 3041
<212> DNA
<213> homo sapiens

<400> 51
atggctttgc ggggcgccgc gggagcgacc gacaccccgg tgtcctcggc cgggggagcc 60

cccggcggct cagcgtcctc gtcgtccacc tcctcgggcg gctcggcctc ggcgggcgcg 120

gggctgtggg ccgcgctcta tgactacgag gctcgcggcg aggacgagct gagcctgcgg 180

cgcggccagc tggtggaggt gctgtcgcag gacgccgccg tgtcgggcga cgagggctgg 240

tgggcaggcc aggtgcagcg gcgcctcggc atcttccccg ccaactacgt ggctccctgc 300

cgcccggccg ccagccccgc gccgccgccc tcgcggccca gctccccggt acacgtcgcc 360

ttcgagcggc tggagctgaa ggagctcatc ggcgctgggg gcttcgggca ggtgtaccgc 420

gccacctggc agggccagga ggtggccgtg aaggcggcgc gccaggaccc ggagcaggac 480

gcggcggcgg ctgccgagag cgtgcggcgc gaggctcggc tcttcgccat gctgcggcac 540

cccaacatca tcgagctgcg cggcgtgtgc ctgcagcagc cgcacctctg cctggtgctg 600

gagttcgccc gcggcggagc gctcaaccga gcgctggccg ctgccaacgc cgccccggac 660

ccgcgcgcgc ccggcccccg ccgcgcgcgc cgcatccctc cgcacgtgct ggtcaactgg 720

gccgtgcaga tagcgcgggg catgctctac ctgcatgagg aggccttcgt gcccatcctg 780

taccgggacc tcaagtccag caacattttg ctacttgaga agatagaaca tgatgacatc 840

tgcaataaaa ctttgaagat tacagatttt gggttggcga gggaatggca caggaccacc 900

aaaatgagca cagcaggcac ctatgcctgg atggcccccg aagtgatcaa gtcttccttg 960

ttttctaagg gaagcgacat ctggagctat ggagtgctgc tgtgggaact gctcaccgga 1020

gaagtcccct atcggggcat tgatggcctc gccgtggctt atggggtagc agtcaataaa 1080

ctcactttgc ccattccatc cacctgccct gagccgtttg ccaagctcat gaaagaatgc 1140

tggcaacaag accctcatat tcgtccatcg tttgccttaa ttctcgaaca gttgactgct 1200

attgaagggg cagtgatgac tgagatgcct caagaatctt ttcattccat gcaagatgac 1260

tggaaactag aaattcaaca aatgtttgat gagttgagaa caaaggaaaa ggagctgcga 1320

tcccgggaag aggagctgac tcgggcggct ctgcagcaga agtctcagga ggagctgcta 1380

aagcggcgtg agcagcagct ggcagagcgc gagatcgacg tgctggagcg ggaacttaac 1440

attctgatat tccagctaaa ccaggagaag cccaaggtaa agaagaggaa gggcaagttt 1500

aagagaagtc gtttaaagct caaagatgga catcgaatca gtttaccttc agatttccag 1560

cacaagataa ccgtgcaggc ctctcccaac ttggacaaac ggcggagcct gaacagcagc 1620

agttccagtc ccccgagcag ccccacaatg atgccccgac tccgagccat acagttgact 1680

tcagatgaaa gcaataaaac ttggggaagg aacacagtct ttcgacaaga agaatttgag 1740

gatgtaaaaa ggaattttaa gaaaaaaggt tgtacctggg gaccaaattc cattcaaatg 1800

aaagatagaa cagattgcaa agaaaggata agacctctct ccgatggcaa cagtccttgg 1860

tcaactatct taataaaaaa tcagaaaacc atgcccttgg cttcattgtt tgtggaccag 1920

ccagggtcct gtgaagagcc aaaactttcc cctgatggat tagaacacag aaaaccaaaa 1980

caaataaaat tgcctagtca ggcctacatt gatctacctc ttgggaaaga tgctcagaga 2040

gagaatcctg cagaagctga aagctgggag gaggcagcct ctgcgaatgc tgccacagtc 2100

tccattgaga tgactcctac gaatagtctg agtagatccc cccagagaaa gaaaacggag 2160

tcagctctgt atgggtgcac cgtccttctg gcatcggtgg ctctgggact ggacctcaga 2220

gagcttcata aagcacaggc tgctgaagaa ccgttgccca aggaagagaa gaagaaacga 2280

gagggaatct tccagcgggc ttccaagtcc cgcagaagcg ccagtcctcc cacaagcctg 2340

ccatccacct gtggggaggc cagcagccca ccctccctgc cactgtcaag tgccctgggc 2400

atcctctcca caccttcttt ctccacaaag tgcctgctgc agatggacag tgaagatcca 2460

ctggtggaca gtgcacctgt cacttgtgac tctgagatgc tcactccgga tttttgtccc 2520

actgccccag gaagtggtcg tgagccagcc ctcatgccaa gacttgacac tgattgtagt 2580

gtatcaagaa acttgccgtc ttccttccta cagcagacat gtgggaatgt accttactgt 2640

gcttcttcaa aacatagacc gtcacatcac agacggacca tgtctgatgg aaatccgacc 2700

ccaactggtg caactattat ctcagccact ggagcctctg cactgccact ctgcccctca 2760

cctgctcctc acagtcatct gccaagggag gtctcaccca agaagcacag cactgtccac 2820

atcgtgcctc agcgtcgccc tgcctccctg agaagccgct cagatctgcc tcaggcttac 2880

ccacagacag cagtgtctca gctggcacag actgcctgtg tagtgggtcg cccaggacca 2940

catcccaccc aattcctcgc tgccaaggag agaactaaat cccatgtgcc ttcattactg 3000

gatgctgacg tggaaggtca gagcagggac tacactgtgc c 3041

<210> 52
<211> 3041
<212> DNA
<213> homo sapiens

<400> 52
atggctttgc ggggcgccgc gggagcgacc gacaccccgg tgtcctcggc cgggggagcc 60

cccggcggct cagcgtcctc gtcgtccacc tcctcgggcg gctcggcctc ggcgggcgcg 120

gggctgtggg ccgcgctcta tgactacgag gctcgcggcg aggacgagct gagcctgcgg 180

cgcggccagc tggtggaggt gctgtcgcag gacgccgccg tgtcgggcga cgagggctgg 240

tgggcaggcc aggtgcagcg gcgcctcggc atcttccccg ccaactacgt ggctccctgc 300

cgcccggccg ccagccccgc gccgccgccc tcgcggccca gctccccggt acacgtcgcc 360

ttcgagcggc tggagctgaa ggagctcatc ggcgctgggg gcttcgggca ggtgtaccgc 420

gccacctggc agggccagga ggtggccgtg aaggcggcgc gccaggaccc ggagcaggac 480

gcggcggcgg ctgccgagag cgtgcggcgc gaggctcggc tcttcgccat gctgcggcac 540

cccaacatca tcgagctgcg cggcgtgtgc ctgcagcagc cgcacctctg cctggtgctg 600

gagttcgccc gcggcggagc gctcaaccga gcgctggccg ctgccaacgc cgccccggac 660

ccgcgcgcgc ccggcccccg ccgcgcgcgc cgcatccctc cgcacgtgct ggtcaactgg 720

gccgtgcaga tagcgcgggg catgctctac ctgcatgagg aggccttcgt gcccatcctg 780

cagcgggacc tcaagtccag caacattttg ctacttgaga agatagaaca tgatgacatc 840

tgcaataaaa ctttgaagat tacagatttt gggttggcga gggaatggca caggaccacc 900

aaaatgagca cagcaggcac ctatgcctgg atggcccccg aagtgatcaa gtcttccttg 960

ttttctaagg gaagcgacat ctggagctat ggagtgctgc tgtgggaact gctcaccgga 1020

gaagtcccct atcggggcat tgatggcctc gccgtggctt atggggtagc agtcaataaa 1080

ctcactttgc ccattccatc cacctgccct gagccgtttg ccaagctcat gaaagaatgc 1140

tggcaacaag accctcatat tcgtccatcg tttgccttaa ttctcgaaca gttgactgct 1200

attgaagggg cagtgatgac tgagatgcct caagaatctt ttcattccat gcaagatgac 1260

tggaaactag aaattcaaca aatgtttgat gagttgagaa caaaggaaaa ggagctgcga 1320

tcccgggaag aggagctgac tcgggcggct ctgcagcaga agtctcagga ggagctgcta 1380

aagcggcgtg agcagcagct ggcagagcgc gagatcgacg tgctggagcg ggaacttaac 1440

attctgatat tccagctaaa ccaggagaag cccaaggtaa agaagaggaa gggcaagttt 1500

aagagaagtc gtttaaagct caaagatgga catcgaatca gtttaccttc agatttccag 1560

cacaagataa ccgtgcaggc ctctcccaac ttggacaaac ggcggagcct gaacagcagc 1620

agttccagtc ccccgagcag ccccacaatg atgccccgac tccgagccat acagttgact 1680

tcagatgaaa gcaataaaac ttggggaagg aacacagtct ttcgacaaga agaatttgag 1740

gatgtaaaaa ggaattttaa gaaaaaaggt tgtacctggg gaccaaattc cattcaaatg 1800

aaagatagaa cagattgcaa agaaaggata agacctctct ccgatggcaa cagtccttgg 1860

tcaactatct taataaaaaa tcagaaaacc atgcccttgg cttcattgtt tgtggaccag 1920

ccagggtcct gtgaagagcc aaaactttcc cctgatggat tagaacacag aaaaccaaaa 1980

caaataaaat tgcctagtca ggcctacatt gatctacctc ttgggaaaga tgctcagaga 2040

gagaatcctg cagaagctga aagctgggag gaggcagcct ctgcgaatgc tgccacagtc 2100

tccattgaga tgactcctac gaatagtctg agtagatccc cccagagaaa gaaaacggag 2160

tcagctctgt atgggtgcac cgtccttctg gcatcggtgg ctctgggact ggacctcaga 2220

gagcttcata aagcacaggc tgctgaagaa ccgttgccca aggaagagaa gaagaaacga 2280

gagggaatct tccagcgggc ttccaagtcc cgcagaagcg ccagtcctcc cacaagcctg 2340

ccatccacct gtggggaggc cagcagccca ccctccctgc cactgtcaag tgccctgggc 2400

atcctctcca caccttcttt ctccacaaag tgcctgctgc agatggacag tgaagatcca 2460

ctggtggaca gtgcacctgt cacttgtgac tctgagatgc tcactccgga tttttgtccc 2520

actgccccag gaagtggtcg tgagccagcc ctcatgccaa gacttgacac tgattgtagt 2580

gtatcaagaa acttgccgtc ttccttccta cagcagacat gtgggaatgt accttactgt 2640

gcttcttcaa aacatagacc gtcacatcac agacggacca tgtctgatgg aaatccgacc 2700

ccaactggtg caactattat ctcagccact ggagcctctg cactgccact ctgcccctca 2760

cctgctcctc acagtcatct gccaagggag gtctcaccca agaagcacag cactgtccac 2820

atcgtgcctc agcgtcgccc tgcctccctg agaagccgct cagatctgcc tcaggcttac 2880

ccacagacag cagtgtctca gctggcacag actgcctgtg tagtgggtcg cccaggacca 2940

catcccaccc aattcctcgc tgccaaggag agaactaaat cccatgtgcc ttcattactg 3000

gatgctgacg tggaaggtca gagcagggac tacactgtgc c 3041

<210> 53
<211> 3041
<212> DNA
<213> homo sapiens

<400> 53
atggctttgc ggggcgccgc gggagcgacc gacaccccgg tgtcctcggc cgggggagcc 60

cccggcggct cagcgtcctc gtcgtccacc tcctcgggcg gctcggcctc ggcgggcgcg 120

gggctgtggg ccgcgctcta tgactacgag gctcgcggcg aggacgagct gagcctgcgg 180
cgcggccagc tggtggaggt gctgtcgcag gacgccgccg tgtcgggcga cgagggctgg 240
tgggcaggcc aggtgcagcg gcgcctcggc atcttccccg ccaactacgt ggctccctgc 300
cgcccggccg ccagccccgc gccgccgccc tcgcggccca gctccccggt acacgtcgcc 360
ttcgagcggc tggagctgaa ggagctcatc ggcgctgggg gcttcgggca ggtgtaccgc 420
gccacctggc agggccagga ggtggccgtg aaggcggcgc gccaggaccc ggagcaggac 480
gcggcggcgg ctgccgagag cgtgcggcgc gaggctcggc tcttcgccat gctgcggcac 540
cccaacatca tcgagctgcg cggcgtgtgc ctgcagcagc cgcacctctg cctggtgctg 600
gagttcgccc gcggcggagc gctcaaccga gcgctggccg ctgccaacgc cgccccggac 660
ccgcgcgcgc ccggcccccg ccgcgcgcgc cgcatccctc cgcacgtgct ggtcaactgg 720
gccgtgcaga tagcgcgggg catgctctac ctgcatgagg aggccttcgt gcccatcctg 780
caccgggacc tcaagtccag caacattttg ctacttgaga agatagaaca tgatgacatc 840
tgcaataaaa ctttgaagat tacagatttt gggttggaga gggaatggca caggaccacc 900
aaaatgagca cagcaggcac ctatgcctgg atggcccccg aagtgatcaa gtcttccttg 960
ttttctaagg gaagcgacat ctggagctat ggagtgctgc tgtgggaact gctcaccgga 1020
gaagtcccct atcggggcat tgatggcctc gccgtggctt atggggtagc agtcaataaa 1080
ctcactttgc ccattccatc cacctgccct gagccgtttg ccaagctcat gaaagaatgc 1140
tggcaacaag accctcatat tcgtccatcg tttgccttaa ttctcgaaca gttgactgct 1200
attgaagggg cagtgatgac tgagatgcct caagaatctt ttcattccat gcaagatgac 1260
tggaaactag aaattcaaca aatgtttgat gagttgagaa caaaggaaaa ggagctgcga 1320
tcccgggaag aggagctgac tcgggcggct ctgcagcaga agtctcagga ggagctgcta 1380
aagcggcgtg agcagcagct ggcagagcgc gagatcgacg tgctggagcg ggaacttaac 1440
attctgatat tccagctaaa ccaggagaag cccaaggtaa agaagaggaa gggcaagttt 1500
aagagaagtc gtttaaagct caaagatgga catcgaatca gtttaccttc agatttccag 1560
cacaagataa ccgtgcaggc ctctcccaac ttggacaaac ggcggagcct gaacagcagc 1620
agttccagtc ccccgagcag ccccacaatg atgccccgac tccgagccat acagttgact 1680
tcagatgaaa gcaataaaac ttggggaagg aacacagtct ttcgacaaga agaatttgag 1740
gatgtaaaaa ggaattttaa gaaaaaaggt tgtacctggg gaccaaattc cattcaaatg 1800
aaagatagaa cagattgcaa agaaaggata agacctctct ccgatggcaa cagtccttgg 1860
tcaactatct taataaaaaa tcagaaaacc atgcccttgg cttcattgtt tgtggaccag 1920
ccagggtcct gtgaagagcc aaaactttcc cctgatggat tagaacacag aaaaccaaaa 1980
caaataaaat tgcctagtca ggcctacatt gatctacctc ttgggaaaga tgctcagaga 2040
gagaatcctg cagaagctga aagctgggag gaggcagcct ctgcgaatgc tgccacagtc 2100
tccattgaga tgactcctac gaatagtctg agtagatccc cccagagaaa gaaaacggag 2160
tcagctctgt atgggtgcac cgtccttctg gcatcggtgg ctctgggact ggacctcaga 2220
gagcttcata aagcacaggc tgctgaagaa ccgttgccca aggaagagaa gaagaaacga 2280

gagggaatct tccagcgggc ttccaagtcc cgcagaagcg ccagtcctcc cacaagcctg 2340

ccatccacct gtggggaggc cagcagccca ccctccctgc cactgtcaag tgccctgggc 2400

atcctctcca caccttcttt ctccacaaag tgcctgctgc agatggacag tgaagatcca 2460

ctggtggaca gtgcacctgt cacttgtgac tctgagatgc tcactccgga tttttgtccc 2520

actgccccag gaagtggtcg tgagccagcc ctcatgccaa gacttgacac tgattgtagt 2580

gtatcaagaa acttgccgtc ttccttccta cagcagacat gtgggaatgt accttactgt 2640

gcttcttcaa aacatagacc gtcacatcac agacggacca tgtctgatgg aaatccgacc 2700

ccaactggtg caactattat ctcagccact ggagcctctg cactgccact ctgcccctca 2760

cctgctcctc acagtcatct gccaagggag gtctcaccca agaagcacag cactgtccac 2820

atcgtgcctc agcgtcgccc tgcctccctg agaagccgct cagatctgcc tcaggcttac 2880

ccacagacag cagtgtctca gctggcacag actgcctgtg tagtgggtcg cccaggacca 2940

catcccaccc aattcctcgc tgccaaggag agaactaaat cccatgtgcc ttcattactg 3000

gatgctgacg tggaaggtca gagcagggac tacactgtgc c 3041

<210> 54
<211> 3041
<212> DNA
<213> homo sapiens

<400> 54

atggctttgc ggggcgccgc gggagcgacc gacaccccgg tgtcctcggc cgggggagcc 60

cccggcggct cagcgtcctc gtcgtccacc tcctcgggcg gctcggcctc ggcgggcgcg 120

gggctgtggg ccgcgctcta tgactacgag gctcgcggcg aggacgagct gagcctgcgg 180

cgcggccagc tggtggaggt gctgtcgcag gacgccgccg tgtcgggcga cgagggctgg 240

tgggcaggcc aggtgcagcg gcgcctcggc atcttccccg ccaactacgt ggctccctgc 300

cgcccggccg ccagccccgc gccgccgccc tcgcggccca gctccccggt acacgtcgcc 360

ttcgagcggc tggagctgaa ggagctcatc ggcgctgggg gcttcgggca ggtgtaccgc 420

gccacctggc agggccagga ggtggccgtg aaggcggcgc gccaggaccc ggagcaggac 480

gcggcggcgg ctgccgagag cgtgcggcgc gaggctcggc tcttcgccat gctgcggcac 540

cccaacatca tcgagctgcg cggcgtgtgc ctgcagcagc cgcacctctg cctggtgctg 600

gagttcgccc gcggcggagc gctcaaccga gcgctggccg ctgccaacgc cgccccggac 660

ccgcgcgcgc ccggcccccg ccgcgcgcgc cgcatccctc cgcacgtgct ggtcaactgg 720

gccgtgcaga tagcgcgggg catgctctac ctgcatgagg aggccttcgt gcccatcctg 780

caccgggacc tcaagtccag caacattttg ctacttgaga agatagaaca tgatgacatc 840

tgcaataaaa ctttgaagat tacagatttt gggttggcga gggaatgaca caggaccacc 900

aaaatgagca cagcaggcac ctatgcctgg atggcccccg aagtgatcaa gtcttccttg 960

ttttctaagg gaagcgacat ctggagctat ggagtgctgc tgtgggaact gctcaccgga 1020

gaagtcccct atcggggcat tgatggcctc gccgtggctt atggggtagc agtcaataaa 1080

ctcactttgc ccattccatc cacctgccct gagccgtttg ccaagctcat gaaagaatgc 1140

tggcaacaag accctcatat tcgtccatcg tttgccttaa ttctcgaaca gttgactgct 1200

attgaagggg cagtgatgac tgagatgcct caagaatctt ttcattccat gcaagatgac 1260

tggaaactag aaattcaaca aatgtttgat gagttgagaa caaaggaaaa ggagctgcga 1320

tcccgggaag aggagctgac tcgggcggct ctgcagcaga agtctcagga ggagctgcta 1380

aagcggcgtg agcagcagct ggcagagcgc gagatcgacg tgctggagcg ggaacttaac 1440

attctgatat tccagctaaa ccaggagaag cccaaggtaa agaagaggaa gggcaagttt 1500

aagagaagtc gtttaaagct caaagatgga catcgaatca gtttaccttc agatttccag 1560

cacaagataa ccgtgcaggc ctctcccaac ttggacaaac ggcggagcct gaacagcagc 1620

agttccagtc ccccgagcag ccccacaatg atgccccgac tccgagccat acagttgact 1680

tcagatgaaa gcaataaaac ttggggaagg aacacagtct ttcgacaaga agaatttgag 1740

gatgtaaaaa ggaattttaa gaaaaaaggt tgtacctggg gaccaaattc cattcaaatg 1800

aaagatagaa cagattgcaa agaaaggata agacctctct ccgatggcaa cagtccttgg 1860

tcaactatct taataaaaaa tcagaaaacc atgcccttgg cttcattgtt tgtggaccag 1920

ccagggtcct gtgaagagcc aaaactttcc cctgatggat tagaacacag aaaaccaaaa 1980

caaataaaat tgcctagtca ggcctacatt gatctacctc ttgggaaaga tgctcagaga 2040

gagaatcctg cagaagctga aagctgggag gaggcagcct ctgcgaatgc tgccacagtc 2100

tccattgaga tgactcctac gaatagtctg agtagatccc cccagagaaa gaaaacggag 2160

tcagctctgt atgggtgcac cgtccttctg gcatcggtgg ctctgggact ggacctcaga 2220

gagcttcata aagcacaggc tgctgaagaa ccgttgccca aggaagagaa gaagaaacga 2280

gagggaatct tccagcgggc ttccaagtcc cgcagaagcg ccagtcctcc cacaagcctg 2340

ccatccacct gtggggaggc cagcagccca ccctccctgc cactgtcaag tgccctgggc 2400

atcctctcca caccttcttt ctccacaaag tgcctgctgc agatggacag tgaagatcca 2460

ctggtggaca gtgcacctgt cacttgtgac tctgagatgc tcactccgga tttttgtccc 2520

actgccccag gaagtggtcg tgagccagcc ctcatgccaa gacttgacac tgattgtagt 2580

gtatcaagaa acttgccgtc ttccttccta cagcagacat gtgggaatgt accttactgt 2640

gcttcttcaa aacatagacc gtcacatcac agacggacca tgtctgatgg aaatccgacc 2700

ccaactggtg caactattat ctcagccact ggagcctctg cactgccact ctgcccctca 2760

cctgctcctc acagtcatct gccaagggag gtctcaccca agaagcacag cactgtccac 2820

atcgtgcctc agcgtcgccc tgcctccctg agaagccgct cagatctgcc tcaggcttac 2880

ccacagacag cagtgtctca gctggcacag actgcctgtg tagtgggtcg cccaggacca 2940

catcccaccc aattcctcgc tgccaaggag agaactaaat cccatgtgcc ttcattactg 3000

gatgctgacg tggaaggtca gagcagggac tacactgtgc c 3041

<210> 55
<211> 3041
<212> DNA
<213> homo sapiens

<400> 55
atggctttgc ggggcgccgc gggagcgacc gacaccccgg tgtcctcggc cgggggagcc 60
cccggcggct cagcgtcctc gtcgtccacc tcctcgggcg gctcggcctc ggcgggcgcg 120
gggctgtggg ccgcgctcta tgactacgag gctcgcggcg aggacgagct gagcctgcgg 180
cgcggccagc tggtggaggt gctgtcgcag gacgccgccg tgtcgggcga cgagggctgg 240
tgggcaggcc aggtgcagcg gcgcctcggc atcttccccg ccaactacgt ggctccctgc 300
cgcccggccg ccagccccgc gccgccgccc tcgcggccca gctccccggt acacgtcgcc 360
ttcgagcggc tggagctgaa ggagctcatc ggcgctgggg gcttcgggca ggtgtaccgc 420
gccacctggc agggccagga ggtggccgtg aaggcggcgc gccaggaccc ggagcaggac 480
gcggcggcgg ctgccgagag cgtgcggcgc gaggctcggc tcttcgccat gctgcggcac 540
cccaacatca tcgagctgcg cggcgtgtgc ctgcagcagc cgcacctctg cctggtgctg 600
gagttcgccc gcggcggagc gctcaaccga gcgctggccg ctgccaacgc cgccccggac 660
ccgcgcgcgc ccggcccccg ccgcgcgcgc cgcatccctc cgcacgtgct ggtcaactgg 720
gccgtgcaga tagcgcgggg catgctctac ctgcatgagg aggccttcgt gcccatcctg 780
caccgggacc tcaagtccag caacattttg ctacttgaga agatagaaca tgatgacatc 840
tgcaataaaa ctttgaagat tacagatttt gggttggcga gggaatggca caggaccacc 900
aaaatgagca cagcaggcac ctatgcctgg atggcccccg aagtgatcaa gtcttccttg 960
ttttctaagg gaagcgacat ctggagctat ggagtgctgc tgtgggaact gctcaccgga 1020
gaagtcccct atcggggcat tgatggcctc gccgtggctt atggggtagc agtcaataaa 1080
ctcactttgc ccattccatc cacctgccct gagccgtttg ccaagctcat gaaagaatgc 1140
tggcaacaag accctcatat tcgtccatcg tttgccttaa ttctcgaaca gttgactgct 1200
attgaagggg cagtgatgac tgagatgcct caagaatctt ttcattccat gcaagatgac 1260
tggaaactag aaattcaaca aatgtttgat gagttgagaa caaaggaaaa ggagctgcga 1320
tcccgggaag aggagctgac tcgggcggct ctgcagcaga agtctcagga ggagctgcta 1380
aagcggcgtg agcagcagct ggcagagcgc gagatcgacg tgctggagcg ggaacttaac 1440
attctgatat tccagctaaa ccaggagaag cccaaggtaa agaagaggaa gggcaagttt 1500
aagagaagtc gtttaaagct caaagatgga catcgaatca gtttaccttc agatttccag 1560
cacaagataa ccgtgcaggc ctctcccaac ttggacaaac ggcggagcct gaacagcagc 1620
agttccagtc ccccgagcag ccccacaatg atgccctgac tccgagccat acagttgact 1680
tcagatgaaa gcaataaaac ttggggaagg aacacagtct ttcgacaaga agaatttgag 1740
gatgtaaaaa ggaattttaa gaaaaaaggt tgtacctggg gaccaaattc cattcaaatg 1800
aaagatagaa cagattgcaa agaaaggata agacctctct ccgatggcaa cagtccttgg 1860
tcaactatct taataaaaaa tcagaaaacc atgcccttgg cttcattgtt tgtggaccag 1920
ccagggtcct gtgaagagcc aaaactttcc cctgatggat tagaacacag aaaaccaaaa 1980
caaataaaat tgcctagtca ggcctacatt gatctacctc ttgggaaaga tgctcagaga 2040
gagaatcctg cagaagctga aagctgggag gaggcagcct ctgcgaatgc tgccacagtc 2100

tccattgaga tgactcctac gaatagtctg agtagatccc cccagagaaa gaaaacggag 2160

tcagctctgt atgggtgcac cgtccttctg gcatcggtgg ctctgggact ggacctcaga 2220

gagcttcata aagcacaggc tgctgaagaa ccgttgccca aggaagagaa gaagaaacga 2280

gagggaatct tccagcgggc ttccaagtcc cgcagaagcg ccagtcctcc cacaagcctg 2340

ccatccacct gtggggaggc cagcagccca ccctccctgc cactgtcaag tgccctgggc 2400

atcctctcca caccttcttt ctccacaaag tgcctgctgc agatggacag tgaagatcca 2460

ctggtggaca gtgcacctgt cacttgtgac tctgagatgc tcactccgga tttttgtccc 2520

actgccccag gaagtggtcg tgagccagcc ctcatgccaa gacttgacac tgattgtagt 2580

gtatcaagaa acttgccgtc ttccttccta cagcagacat gtgggaatgt accttactgt 2640

gcttcttcaa aacatagacc gtcacatcac agacggacca tgtctgatgg aaatccgacc 2700

ccaactggtg caactattat ctcagccact ggagcctctg cactgccact ctgcccctca 2760

cctgctcctc acagtcatct gccaagggag gtctcaccca agaagcacag cactgtccac 2820

atcgtgcctc agcgtcgccc tgcctccctg agaagccgct cagatctgcc tcaggcttac 2880

ccacagacag cagtgtctca gctggcacag actgcctgtg tagtgggtcg cccaggacca 2940

catcccaccc aattcctcgc tgccaaggag agaactaaat cccatgtgcc ttcattactg 3000

gatgctgacg tggaaggtca gagcagggac tacactgtgc c 3041

<210> 56
<211> 3041
<212> DNA
<213> homo sapiens

<400> 56
atggctttgc ggggcgccgc gggagcgacc gacaccccgg tgtcctcggc cgggggagcc 60

cccggcggct cagcgtcctc gtcgtccacc tcctcgggcg gctcggcctc ggcgggcgcg 120

gggctgtggg ccgcgctcta tgactacgag gctcgcggcg aggacgagct gagcctgcgg 180

cgcggccagc tggtggaggt gctgtcgcag gacgccgccg tgtcgggcga cgagggctgg 240

tgggcaggcc aggtgcagcg gcgcctcggc atcttccccg ccaactacgt ggctccctgc 300

cgcccggccg ccagccccgc gccgccgccc tcgcggccca gctccccggt acacgtcgcc 360

ttcgagcggc tggagctgaa ggagctcatc ggcgctgggg gcttcgggca ggtgtaccgc 420

gccacctggc agggccagga ggtggccgtg aaggcggcgc gccaggaccc ggagcaggac 480

gcggcggcgg ctgccgagag cgtgcggcgc gaggctcggc tcttcgccat gctgcggcac 540

cccaacatca tcgagctgcg cggcgtgtgc ctgcagcagc cgcacctctg cctggtgctg 600

gagttcgccc gcggcggagc gctcaaccga gcgctggccg ctgccaacgc cgccccggac 660

ccgcgcgcgc ccggcccccg ccgcgcgcgc cgcatccctc cgcacgtgct ggtcaactgg 720

gccgtgcaga tagcgcgggg catgctctac ctgcatgagg aggccttcgt gcccatcctg 780

caccgggacc tcaagtccag caacattttg ctacttgaga agatagaaca tgatgacatc 840

tgcaataaaa ctttgaagat tacagatttt gggttggcga gggaatggca caggaccacc 900

aaaatgagca cagcaggcac ctatgcctgg atggcccccg aagtgatcaa gtcttccttg 960

ttttctaagg gaagcgacat ctggagctat ggagtgctgc tgtgggaact gctcaccgga 1020

gaagtcccct atcggggcat tgatggcctc gccgtggctt atggggtagc agtcaataaa 1080

ctcactttgc ccattccatc cacctgccct gagccgtttg ccaagctcat gaaagaatgc 1140

tggcaacaag accctcatat tcgtccatcg tttgccttaa ttctcgaaca gttgactgct 1200

attgaagggg cagtgatgac tgagatgcct caagaatctt ttcattccat gcaagatgac 1260

tggaaactag aaattcaaca aatgtttgat gagttgagaa caaaggaaaa ggagctgcga 1320

tcccgggaag aggagctgac tcgggcggct ctgcagcaga agtctcagga ggagctgcta 1380

aagcggcgtg agcagcagct ggcagagcgc gagatcgacg tgctggagcg ggaacttaac 1440

attctgatat tccagctaaa ccaggagaag cccaaggtaa agaagaggaa gggcaagttt 1500

aagagaagtc gtttaaagct caaagatgga catcgaatca gtttaccttc agatttccag 1560

cacaagataa ccgtgcaggc ctctcccaac ttggacaaac ggcggagcct gaacagcagc 1620

agttccagtc ccccgagcag ccccacaatg atgccccgac tctgagccat acagttgact 1680

tcagatgaaa gcaataaaac ttggggaagg aacacagtct ttcgacaaga agaatttgag 1740

gatgtaaaaa ggaattttaa gaaaaaaggt tgtacctggg gaccaaattc cattcaaatg 1800

aaagatagaa cagattgcaa agaaaggata agacctctct ccgatggcaa cagtccttgg 1860

tcaactatct taataaaaaa tcagaaaacc atgcccttgg cttcattgtt tgtggaccag 1920

ccagggtcct gtgaagagcc aaaactttcc cctgatggat tagaacacag aaaaccaaaa 1980

caaataaaat tgcctagtca ggcctacatt gatctacctc ttgggaaaga tgctcagaga 2040

gagaatcctg cagaagctga aagctgggag gaggcagcct ctgcgaatgc tgccacagtc 2100

tccattgaga tgactcctac gaatagtctg agtagatccc cccagagaaa gaaaacggag 2160

tcagctctgt atgggtgcac cgtccttctg gcatcggtgg ctctgggact ggacctcaga 2220

gagcttcata aagcacaggc tgctgaagaa ccgttgccca aggaagagaa gaagaaacga 2280

gagggaatct tccagcgggc ttccaagtcc cgcagaagcg ccagtcctcc cacaagcctg 2340

ccatccacct gtggggaggc cagcagccca ccctccctgc cactgtcaag tgccctgggc 2400

atcctctcca caccttcttt ctccacaaag tgcctgctgc agatggacag tgaagatcca 2460

ctggtggaca gtgcacctgt cacttgtgac tctgagatgc tcactccgga tttttgtccc 2520

actgccccag gaagtggtcg tgagccagcc ctcatgccaa gacttgacac tgattgtagt 2580

gtatcaagaa acttgccgtc ttccttccta cagcagacat gtgggaatgt accttactgt 2640

gcttcttcaa aacatagacc gtcacatcac agacggacca tgtctgatgg aaatccgacc 2700

ccaactggtg caactattat ctcagccact ggagcctctg cactgccact ctgcccctca 2760

cctgctcctc acagtcatct gccaagggag gtctcaccca agaagcacag cactgtccac 2820

atcgtgcctc agcgtcgccc tgcctccctg agaagccgct cagatctgcc tcaggcttac 2880

ccacagacag cagtgtctca gctggcacag actgcctgtg tagtgggtcg cccaggacca 2940

catcccaccc aattcctcgc tgccaaggag agaactaaat cccatgtgcc ttcattactg 3000

gatgctgacg tggaaggtca gagcagggac tacactgtgc c 3041

<210> 57
<211> 3041
<212> DNA
<213> homo sapiens

<400> 57
atggctttgc ggggcgccgc gggagcgacc gacaccccgg tgtcctcggc cgggggagcc 60

cccggcggct cagcgtcctc gtcgtccacc tcctcgggcg gctcggcctc ggcgggcgcg 120

gggctgtggg ccgcgctcta tgactacgag gctcgcggcg aggacgagct gagcctgcgg 180

cgcggccagc tggtggaggt gctgtcgcag gacgccgccg tgtcgggcga cgagggctgg 240

tgggcaggcc aggtgcagcg gcgcctcggc atcttccccg ccaactacgt ggctccctgc 300

cgcccggccg ccagccccgc gccgccgccc tcgcggccca gctccccggt acacgtcgcc 360

ttcgagcggc tggagctgaa ggagctcatc ggcgctgggg gcttcgggca ggtgtaccgc 420

gccacctggc agggccagga ggtggccgtg aaggcggcgc gccaggaccc ggagcaggac 480

gcggcggcgg ctgccgagag cgtgcggcgc gaggctcggc tcttcgccat gctgcggcac 540

cccaacatca tcgagctgcg cggcgtgtgc ctgcagcagc cgcacctctg cctggtgctg 600

gagttcgccc gcggcggagc gctcaaccga gcgctggccg ctgccaacgc cgccccggac 660

ccgcgcgcgc ccggcccccg ccgcgcgcgc cgcatccctc cgcacgtgct ggtcaactgg 720

gccgtgcaga tagcgcgggg catgctctac ctgcatgagg aggccttcgt gcccatcctg 780

caccgggacc tcaagtccag caacattttg ctacttgaga agatagaaca tgatgacatc 840

tgcaataaaa ctttgaagat tacagatttt gggttggcga gggaatggca caggaccacc 900

aaaatgagca cagcaggcac ctatgcctgg atggcccccg aagtgatcaa gtcttccttg 960

ttttctaagg gaagcgacat ctggagctat ggagtgctgc tgtgggaact gctcaccgga 1020

gaagtcccct atcggggcat tgatggcctc gccgtggctt atggggtagc agtcaataaa 1080

ctcactttgc ccattccatc cacctgccct gagccgtttg ccaagctcat gaaagaatgc 1140

tggcaacaag accctcatat tcgtccatcg tttgccttaa ttctcgaaca gttgactgct 1200

attgaagggg cagtgatgac tgagatgcct caagaatctt ttcattccat gcaagatgac 1260

tggaaactag aaattcaaca aatgtttgat gagttgagaa caaaggaaaa ggagctgcga 1320

tcccgggaag aggagctgac tcgggcggct ctgcagcaga agtctcagga ggagctgcta 1380

aagcggcgtg agcagcagct ggcagagcgc gagatcgacg tgctggagcg ggaacttaac 1440

attctgatat tccagctaaa ccaggagaag cccaaggtaa agaagaggaa gggcaagttt 1500

aagagaagtc gtttaaagct caaagatgga catcgaatca gtttaccttc agatttccag 1560

cacaagataa ccgtgcaggc ctctcccaac ttggacaaac ggcggagcct gaacagcagc 1620

agttccagtc ccccgagcag ccccacaatg atgccccgac tccgagccat acagttgact 1680

tcagatgaaa gcaataaaac ttggggaagg aacacagtct ttcgacaaga agaatttgag 1740

gatgtaaaaa ggaattttaa gaaaaaaggt tgtacctggg gaccaatttc cattcaaatg 1800

aaagatagaa cagattgcaa agaaaggata agacctctct ccgatggcaa cagtccttgg 1860

tcaactatct taataaaaaa tcagaaaacc atgcccttgg cttcattgtt tgtggaccag 1920

ccagggtcct gtgaagagcc aaaactttcc cctgatggat tagaacacag aaaaccaaaa 1980

caaataaaat tgcctagtca ggcctacatt gatctacctc ttgggaaaga tgctcagaga 2040

gagaatcctg cagaagctga aagctgggag gaggcagcct ctgcgaatgc tgccacagtc 2100

tccattgaga tgactcctac gaatagtctg agtagatccc cccagagaaa gaaaacggag 2160

tcagctctgt atgggtgcac cgtccttctg gcatcggtgg ctctgggact ggacctcaga 2220

gagcttcata aagcacaggc tgctgaagaa ccgttgccca aggaagagaa gaagaaacga 2280

gagggaatct tccagcgggc ttccaagtcc cgcagaagcg ccagtcctcc cacaagcctg 2340

ccatccacct gtggggaggc agcagccca ccctccctgc cactgtcaag tgccctgggc 2400

atcctctcca caccttcttt ctccacaaag tgcctgctgc agatggacag tgaagatcca 2460

ctggtggaca gtgcacctgt cacttgtgac tctgagatgc tcactccgga tttttgtccc 2520

actgccccag gaagtggtcg tgagccagcc ctcatgccaa gacttgacac tgattgtagt 2580

gtatcaagaa acttgccgtc ttccttccta cagcagacat gtgggaatgt accttactgt 2640

gcttcttcaa aacatagacc gtcacatcac agacggacca tgtctgatgg aaatccgacc 2700

ccaactggtg caactattat ctcagccact ggagcctctg cactgccact ctgcccctca 2760

cctgctcctc acagtcatct gccaagggag gtctcaccca agaagcacag cactgtccac 2820

atcgtgcctc agcgtcgccc tgcctccctg agaagccgct cagatctgcc tcaggcttac 2880

ccacagacag cagtgtctca gctggcacag actgcctgtg tagtgggtcg cccaggacca 2940

catcccaccc aattcctcgc tgccaaggag agaactaaat cccatgtgcc ttcattactg 3000

gatgctgacg tggaaggtca gagcagggac tacactgtgc c 3041

<210> 58
<211> 3041
<212> DNA
<213> homo sapiens

<400> 58
atggctttgc ggggcgccgc gggagcgacc gacaccccgg tgtcctcggc cgggggagcc 60

cccggcggct cagcgtcctc gtcgtccacc tcctcgggcg gctcggcctc ggcgggcgcg 120

gggctgtggg ccgcgctcta tgactacgag gctcgcggcg aggacgagct gagcctgcgg 180

cgcggccagc tggtggaggt gctgtcgcag gacgccgccg tgtcgggcga cgagggctgg 240

tgggcaggcc aggtgcagcg gcgcctcggc atcttccccg ccaactacgt ggctccctgc 300

cgcccggccg ccagccccgc gccgccgccc tcgcggccca gctccccggt acacgtcgcc 360

ttcgagcggc tggagctgaa ggagctcatc ggcgctgggg gcttcgggca ggtgtaccgc 420

gccacctggc agggccagga ggtggccgtg aaggcggcgc gccaggaccc ggagcaggac 480

gcggcggcgg ctgccgagag cgtgcggcgc gaggctcggc tcttcgccat gctgcggcac 540

cccaacatca tcgagctgcg cggcgtgtgc ctgcagcagc cgcacctctg cctggtgctg 600

gagttcgccc gcggcggagc gctcaaccga gcgctggccg ctgccaacgc cgccccggac 660

ccgcgcgcgc ccggcccccg ccgcgcgcgc cgcatccctc cgcacgtgct ggtcaactgg 720

gccgtgcaga tagcgcgggg catgctctac ctgcatgagg aggccttcgt gcccatcctg 780

caccgggacc tcaagtccag caacattttg ctacttgaga agatagaaca tgatgacatc 840

tgcaataaaa ctttgaagat tacagatttt gggttggcga gggaatggca caggaccacc 900

aaaatgagca cagcaggcac ctatgcctgg atggcccccg aagtgatcaa gtcttccttg 960

ttttctaagg gaagcgacat ctggagctat ggagtgctgc tgtgggaact gctcaccgga 1020

gaagtcccct atcggggcat tgatggcctc gccgtggctt atggggtagc agtcaataaa 1080

ctcactttgc ccattccatc cacctgccct gagccgtttg ccaagctcat gaaagaatgc 1140

tggcaacaag accctcatat tcgtccatcg tttgccttaa ttctcgaaca gttgactgct 1200

attgaagggg cagtgatgac tgagatgcct caagaatctt ttcattccat gcaagatgac 1260

tggaaactag aaattcaaca aatgtttgat gagttgagaa caaaggaaaa ggagctgcga 1320

tcccgggaag aggagctgac tcgggcggct ctgcagcaga agtctcagga ggagctgcta 1380

aagcggcgtg agcagcagct ggcagagcgc gagatcgacg tgctggagcg ggaacttaac 1440

attctgatat tccagctaaa ccaggagaag cccaaggtaa agaagaggaa gggcaagttt 1500

aagagaagtc gtttaaagct caaagatgga catcgaatca gtttaccttc agatttccag 1560

cacaagataa ccgtgcaggc ctctcccaac ttggacaaac ggcggagcct gaacagcagc 1620

agttccagtc ccccgagcag ccccacaatg atgccccgac tccgagccat acagttgact 1680

tcagatgaaa gcaataaaac ttggggaagg aacacagtct ttcgacaaga agaatttgag 1740

gatgtaaaaa ggaattttaa gaaaaaaggt tgtacctggg gaccaaattc cattcaaatg 1800

aaagatagaa cagattgcaa agaaaggata agacctctct ccgatggcaa cagtccttgg 1860

tcaactatct taataaaaaa tcaggaaacc atgcccttgg cttcattgtt tgtggaccag 1920

ccagggtcct gtgaagagcc aaaactttcc cctgatggat tagaacacag aaaaccaaaa 1980

caaataaaat tgcctagtca ggcctacatt gatctacctc ttgggaaaga tgctcagaga 2040

gagaatcctg cagaagctga aagctgggag gaggcagcct ctgcgaatgc tgccacagtc 2100

tccattgaga tgactcctac gaatagtctg agtagatccc cccagagaaa gaaaacggag 2160

tcagctctgt atgggtgcac cgtccttctg gcatcggtgg ctctgggact ggacctcaga 2220

gagcttcata aagcacaggc tgctgaagaa ccgttgccca aggaagagaa gaagaaacga 2280

gagggaatct tccagcgggc ttccaagtcc cgcagaagcg ccagtcctcc cacaagcctg 2340

ccatccacct gtggggaggc cagcagccca ccctccctgc cactgtcaag tgccctgggc 2400

atcctctcca caccttcttt ctccacaaag tgcctgctgc agatggacag tgaagatcca 2460

ctggtggaca gtgcacctgt cacttgtgac tctgagatgc tcactccgga tttttgtccc 2520

actgccccag gaagtggtcg tgagccagcc ctcatgccaa gacttgacac tgattgtagt 2580

gtatcaagaa acttgccgtc ttccttccta cagcagacat gtgggaatgt accttactgt 2640

gcttcttcaa aacatagacc gtcacatcac agacggacca tgtctgatgg aaatccgacc 2700

ccaactggtg caactattat ctcagccact ggagcctctg cactgccact ctgccctca 2760

cctgctcctc acagtcatct gccaagggag gtctcaccca agaagcacag cactgtccac 2820

atcgtgcctc agcgtcgccc tgcctccctg agaagccgct cagatctgcc tcaggcttac 2880

ccacagacag cagtgtctca gctggcacag actgcctgtg tagtgggtcg cccaggacca 2940

catcccaccc aattcctcgc tgccaaggag agaactaaat cccatgtgcc ttcattactg 3000

gatgctgacg tggaaggtca gagcagggac tacactgtgc c 3041

<210> 59
<211> 3041
<212> DNA
<213> homo sapiens

<400> 59
atggctttgc ggggcgccgc gggagcgacc gacaccccgg tgtcctcggc cgggggagcc 60
cccggcggct cagcgtcctc gtcgtccacc tcctcgggcg gctcggcctc ggcgggcgcg 120
gggctgtggg ccgcgctcta tgactacgag gctcgcggcg aggacgagct gagcctgcgg 180
cgcggccagc tggtggaggt gctgtcgcag gacgccgccg tgtcgggcga cgagggctgg 240
tgggcaggcc aggtgcagcg gcgcctcggc atcttccccg ccaactacgt ggctccctgc 300
cgcccggccg ccagccccgc gccgccgccc tcgcggccca gctccccggt acacgtcgcc 360
ttcgagcggc tggagctgaa ggagctcatc ggcgctgggg gcttcgggca ggtgtaccgc 420
gccacctggc agggccagga ggtggccgtg aaggcggcgc gccaggaccc ggagcaggac 480
gcggcggcgg ctgccgagag cgtgcggcgc gaggctcggc tcttcgccat gctgcggcac 540
cccaacatca tcgagctgcg cggcgtgtgc ctgcagcagc cgcacctctg cctggtgctg 600
gagttcgccc gcggcggagc gctcaaccga gcgctggccg ctgccaacgc cgccccggac 660
ccgcgcgcgc ccggcccccg ccgcgcgcgc cgcatccctc cgcacgtgct ggtcaactgg 720
gccgtgcaga tagcgcgggg catgctctac ctgcatgagg aggccttcgt gcccatcctg 780
caccgggacc tcaagtccag caacattttg ctacttgaga agatagaaca tgatgacatc 840
tgcaataaaa ctttgaagat tacagatttt gggttggcga gggaatggca caggaccacc 900
aaaatgagca cagcaggcac ctatgcctgg atggcccccg aagtgatcaa gtcttccttg 960
ttttctaagg gaagcgacat ctggagctat ggagtgctgc tgtgggaact gctcaccgga 1020
gaagtcccct atcggggcat tgatggcctc gccgtggctt atggggtagc agtcaataaa 1080
ctcactttgc ccattccatc cacctgccct gagccgtttg ccaagctcat gaaagaatgc 1140
tggcaacaag accctcatat tcgtccatcg tttgccttaa ttctcgaaca gttgactgct 1200
attgaagggg cagtgatgac tgagatgcct caagaatctt ttcattccat gcaagatgac 1260
tggaaactag aaattcaaca aatgtttgat gagttgagaa caaaggaaaa ggagctgcga 1320
tcccgggaag aggagctgac tcgggcggct ctgcagcaga agtctcagga ggagctgcta 1380
aagcggcgtg agcagcagct ggcagagcgc gagatcgacg tgctggagcg ggaacttaac 1440
attctgatat tccagctaaa ccaggagaag cccaaggtaa agaagaggaa gggcaagttt 1500
aagagaagtc gtttaaagct caaagatgga catcgaatca gtttaccttc agatttccag 1560
cacaagataa ccgtgcaggc ctctcccaac ttggacaaac ggcggagcct gaacagcagc 1620
agttccagtc ccccgagcag ccccacaatg atgccccgac tccgagccat acagttgact 1680
tcagatgaaa gcaataaaac ttggggaagg aacacagtct ttcgacaaga agaatttgag 1740
gatgtaaaaa ggaattttaa gaaaaaaggt tgtacctggg gaccaaattc cattcaaatg 1800

aaagatagaa cagattgcaa agaaaggata agacctctct ccgatggcaa cagtccttgg 1860

tcaactatct taataaaaaa tcagaaaacc atgcccttgg cttcattgtt tgtggaccag 1920

ccagggtcct gtgaagagcc aaaactttcc cctgatggat tagaacacag aaaaccaaaa 1980

caaataaaat tgcctagtca ggcctacatt gatctacctc ttgggaaaga tgctcagaga 2040

gagaatcctg cagaagctga aagctgggag gaggcagcct ctgcgaatgc tgccacagtc 2100

tccattgaga tgactcctac gaatagtctg agtagatccc cccagagaaa gaaaacggag 2160

tcagctctgt atgggtgcac cgtccttctg gcatcggtgg ctctgggact ggacctcaga 2220

gagcttcata aagcacaggc tgctgaagaa ccgttgccca aggaagagaa gaagaaacga 2280

gagggaatct tccagcgggc ttccaagtcc cgcagaagcg ccagtcctcc cacaagcctg 2340

ccatccacct gtggggaggc cagcagccca ccctccctgc cactgtcaag tgccctgggc 2400

atcctctcca caccttcttt ctccacaaag tgcctgctgc agatggacag tgaagatcca 2460

ctggtggaca gtgcacctgt cacttgtgac tctgagatgc tcactccgga tttttgtccc 2520

actgcctcag gaagtggtcg tgagccagcc ctcatgccaa gacttgacac tgattgtagt 2580

gtatcaagaa acttgccgtc ttccttccta cagcagacat gtgggaatgt accttactgt 2640

gcttcttcaa aacatagacc gtcacatcac agacggacca tgtctgatgg aaatccgacc 2700

ccaactggtg caactattat ctcagccact ggagcctctg cactgccact ctgcccctca 2760

cctgctcctc acagtcatct gccaagggag gtctcaccca agaagcacag cactgtccac 2820

atcgtgcctc agcgtcgccc tgcctccctg agaagccgct cagatctgcc tcaggcttac 2880

ccacagacag cagtgtctca gctggcacag actgcctgtg tagtgggtcg cccaggacca 2940

catcccaccc aattcctcgc tgccaaggag agaactaaat cccatgtgcc ttcattactg 3000

gatgctgacg tggaaggtca gagcagggac tacactgtgc c 3041

<210> 60
<211> 3041
<212> DNA
<213> homo sapiens

<400> 60
atggctttgc ggggcgccgc gggagcgacc gacaccccgg tgtcctcggc cgggggagcc 60

cccggcggct cagcgtcctc gtcgtccacc tcctcgggcg gctcggcctc ggcgggcgcg 120

gggctgtggg ccgcgctcta tgactacgag gctcgcggcg aggacgagct gagcctgcgg 180

cgcggccagc tggtggaggt gctgtcgcag gacgccgccg tgtcgggcga cgagggctgg 240

tgggcaggcc aggtgcagcg gcgcctcggc atcttccccg ccaactacgt ggctccctgc 300

cgcccggccg ccagccccgc gccgccgccc tcgcggccca gctccccggt acacgtcgcc 360

ttcgagcggc tggagctgaa ggagctcatc ggcgctgggg gcttcgggca ggtgtaccgc 420

gccacctggc agggccagga ggtggccgtg aaggcggcgc gccaggaccc ggagcaggac 480

gcggcggcgg ctgccgagag cgtgcggcgc gaggctcggc tcttcgccat gctgcggcac 540

cccaacatca tcgagctgcg cggcgtgtgc ctgcagcagc cgcacctctg cctggtgctg 600

gagttcgccc gcggcggagc gctcaaccga gcgctggccg ctgccaacgc cgccccggac 660

ccgcgcgcgc ccggcccccg ccgcgcgcgc cgcatccctc cgcacgtgct ggtcaactgg 720

gccgtgcaga tagcgcgggg catgctctac ctgcatgagg aggccttcgt gcccatcctg 780

caccgggacc tcaagtccag caacattttg ctacttgaga agatagaaca tgatgacatc 840

tgcaataaaa ctttgaagat tacagatttt gggttggcga gggaatggca caggaccacc 900

aaaatgagca cagcaggcac ctatgcctgg atggcccccg aagtgatcaa gtcttccttg 960

ttttctaagg gaagcgacat ctggagctat ggagtgctgc tgtgggaact gctcaccgga 1020

gaagtcccct atcggggcat tgatggcctc gccgtggctt atggggtagc agtcaataaa 1080

ctcactttgc ccattccatc cacctgccct gagccgtttg ccaagctcat gaaagaatgc 1140

tggcaacaag accctcatat tcgtccatcg tttgccttaa ttctcgaaca gttgactgct 1200

attgaagggg cagtgatgac tgagatgcct caagaatctt ttcattccat gcaagatgac 1260

tggaaactag aaattcaaca aatgtttgat gagttgagaa caaaggaaaa ggagctgcga 1320

tcccgggaag aggagctgac tcgggcggct ctgcagcaga agtctcagga ggagctgcta 1380

aagcggcgtg agcagcagct ggcagagcgc gagatcgacg tgctggagcg ggaacttaac 1440

attctgatat tccagctaaa ccaggagaag cccaaggtaa agaagaggaa gggcaagttt 1500

aagagaagtc gtttaaagct caaagatgga catcgaatca gtttaccttc agatttccag 1560

cacaagataa ccgtgcaggc ctctcccaac ttggacaaac ggcggagcct gaacagcagc 1620

agttccagtc ccccgagcag ccccacaatg atgccccgac tccgagccat acagttgact 1680

tcagatgaaa gcaataaaac ttggggaagg aacacagtct ttcgacaaga agaatttgag 1740

gatgtaaaaa ggaattttaa gaaaaaaggt tgtacctggg gaccaaattc cattcaaatg 1800

aaagatagaa cagattgcaa agaaaggata agacctctct ccgatggcaa cagtccttgg 1860

tcaactatct taataaaaaa tcagaaaacc atgcccttgg cttcattgtt tgtggaccag 1920

ccagggtcct gtgaagagcc aaaactttcc cctgatggat tagaacacag aaaaccaaaa 1980

caaataaaat tgcctagtca ggcctacatt gatctacctc ttgggaaaga tgctcagaga 2040

gagaatcctg cagaagctga aagctgggag gaggcagcct ctgcgaatgc tgccacagtc 2100

tccattgaga tgactcctac gaatagtctg agtagatccc cccagagaaa gaaaacggag 2160

tcagctctgt atgggtgcac cgtccttctg gcatcggtgg ctctgggact ggacctcaga 2220

gagcttcata aagcacaggc tgctgaagaa ccgttgccca aggaagagaa gaagaaacga 2280

gagggaatct tccagcgggc ttccaagtcc cgcagaagcg ccagtcctcc cacaagcctg 2340

ccatccacct gtggggaggc cagcagccca ccctccctgc cactgtcaag tgccctgggc 2400

atcctctcca caccttcttt ctccacaaag tgcctgctgc agatggacag tgaagatcca 2460

ctggtggaca gtgcacctgt cacttgtgac tctgagatgc tcactccgga tttttgtccc 2520

actgccccag gaagtggtcg tgagccagcc ctcatgccaa gacttgacac tgattgtagt 2580

gtatcaagaa acttgccgtc ttccttccta cagcagacat gtgggaatgt accttactgt 2640

gcttcttcaa aacatagacc gtcacatccc agacggacca tgtctgatgg aaatccgacc 2700

ccaactggtg caactattat ctcagccact ggagcctctg cactgccact ctgcccctca 2760

cctgctcctc acagtcatct gccaagggag gtctcaccca agaagcacag cactgtccac 2820

atcgtgcctc agcgtcgccc tgcctccctg agaagccgct cagatctgcc tcaggcttac 2880

ccacagacag cagtgtctca gctggcacag actgcctgtg tagtgggtcg cccaggacca 2940

catcccaccc aattcctcgc tgccaaggag agaactaaat cccatgtgcc ttcattactg 3000

gatgctgacg tggaaggtca gagcagggac tacactgtgc c 3041

<210> 61
<211> 3041
<212> DNA
<213> homo sapiens

<400> 61
atggctttgc ggggcgccgc gggagcgacc gacaccccgg tgtcctcggc cgggggagcc 60

cccggcggct cagcgtcctc gtcgtccacc tcctcgggcg gctcggcctc ggcgggcgcg 120

gggctgtggg ccgcgctcta tgactacgag gctcgcggcg aggacgagct gagcctgcgg 180

cgcggccagc tggtggaggt gctgtcgcag gacgccgccg tgtcgggcga cgagggctgg 240

tgggcaggcc aggtgcagcg gcgcctcggc atcttccccg ccaactacgt ggctccctgc 300

cgcccggccg ccagccccgc gccgccgccc tcgcggccca gctccccggt acacgtcgcc 360

ttcgagcggc tggagctgaa ggagctcatc ggcgctgggg gcttcgggca ggtgtaccgc 420

gccacctggc agggccagga ggtggccgtg aaggcggcgc gccaggaccc ggagcaggac 480

gcggcggcgg ctgccgagag cgtgcggcgc gaggctcggc tcttcgccat gctgcggcac 540

cccaacatca tcgagctgcg cggcgtgtgc ctgcagcagc cgcacctctg cctggtgctg 600

gagttcgccc gcggcggagc gctcaaccga gcgctggccg ctgccaacgc cgccccggac 660

ccgcgcgcgc ccggcccccg ccgcgcgcgc cgcatccctc cgcacgtgct ggtcaactgg 720

gccgtgcaga tagcgcgggg catgctctac ctgcatgagg aggccttcgt gcccatcctg 780

caccgggacc tcaagtccag caacattttg ctacttgaga agatagaaca tgatgacatc 840

tgcaataaaa ctttgaagat tacagatttt gggttggcga gggaatggca caggaccacc 900

aaaatgagca cagcaggcac ctatgcctgg atggcccccg aagtgatcaa gtcttccttg 960

ttttctaagg gaagcgacat ctggagctat ggagtgctgc tgtgggaact gctcaccgga 1020

gaagtcccct atcggggcat tgatggcctc gccgtggctt atggggtagc agtcaataaa 1080

ctcactttgc ccattccatc cacctgccct gagccgtttg ccaagctcat gaaagaatgc 1140

tggcaacaag accctcatat tcgtccatcg tttgccttaa ttctcgaaca gttgactgct 1200

attgaagggg cagtgatgac tgagatgcct caagaatctt ttcattccat gcaagatgac 1260

tggaaactag aaattcaaca aatgtttgat gagttgagaa caaaggaaaa ggagctgcga 1320

tcccgggaag aggagctgac tcgggcggct ctgcagcaga agtctcagga ggagctgcta 1380

aagcggcgtg agcagcagct ggcagagcgc gagatcgacg tgctggagcg ggaacttaac 1440

attctgatat tccagctaaa ccaggagaag cccaaggtaa agaagaggaa gggcaagttt 1500

aagagaagtc gtttaaagct caaagatgga catcgaatca gtttaccttc agatttccag 1560

cacaagataa ccgtgcaggc ctctcccaac ttggacaaac ggcggagcct gaacagcagc 1620

agttccagtc ccccgagcag ccccacaatg atgccccgac tccgagccat acagttgact 1680

tcagatgaaa gcaataaaac ttggggaagg aacacagtct ttcgacaaga agaatttgag 1740

gatgtaaaaa ggaattttaa gaaaaaaggt tgtacctggg gaccaaattc cattcaaatg 1800

aaagatagaa cagattgcaa agaaaggata agacctctct ccgatggcaa cagtccttgg 1860

tcaactatct taataaaaaa tcagaaaacc atgcccttgg cttcattgtt tgtggaccag 1920

ccagggtcct gtgaagagcc aaaactttcc cctgatggat tagaacacag aaaaccaaaa 1980

caaataaaat tgcctagtca ggcctacatt gatctacctc ttgggaaaga tgctcagaga 2040

gagaatcctg cagaagctga aagctgggag gaggcagcct ctgcgaatgc tgccacagtc 2100

tccattgaga tgactcctac gaatagtctg agtagatccc cccagagaaa gaaaacggag 2160

tcagctctgt atgggtgcac cgtccttctg gcatcggtgg ctctgggact ggacctcaga 2220

gagcttcata aagcacaggc tgctgaagaa ccgttgccca aggaagagaa gaagaaacga 2280

gagggaatct tccagcgggc ttccaagtcc cgcagaagcg ccagtcctcc cacaagcctg 2340

ccatccacct gtggggaggc cagcagccca ccctccctgc cactgtcaag tgccctgggc 2400

atcctctcca caccttcttt ctccacaaag tgcctgctgc agatggacag tgaagatcca 2460

ctggtggaca gtgcacctgt cacttgtgac tctgagatgc tcactccgga tttttgtccc 2520

actgccccag gaagtggtcg tgagccagcc ctcatgccaa gacttgacac tgattgtagt 2580

gtatcaagaa acttgccgtc ttccttccta cagcagacat gtgggaatgt accttactgt 2640

gcttcttcaa aacatagacc gtcacatcac agacggacca cgtctgatgg aaatccgacc 2700

ccaactggtg caactattat ctcagccact ggagcctctg cactgccact ctgcccctca 2760

cctgctcctc acagtcatct gccaagggag gtctcaccca agaagcacag cactgtccac 2820

atcgtgcctc agcgtcgccc tgcctccctg agaagccgct cagatctgcc tcaggcttac 2880

ccacagacag cagtgtctca gctggcacag actgcctgtg tagtgggtcg cccaggacca 2940

catcccaccc aattcctcgc tgccaaggag agaactaaat cccatgtgcc ttcattactg 3000

gatgctgacg tggaaggtca gagcagggac tacactgtgc c 3041

Bibliografy

1. A. Bardelli *et al.*, *Science (New York, N.Y* **300**, 949 (May 9, 2003).
2. H. Davies *et al.*, *Nature* **417**, 949 (Jun 27, 2002).
3. T. Maeda *et al.*, *Nature* **433**, 278 (Jan 20, 2005).
4. S. Shirasawa, M. Furuse, N. Yokoyama, T. Sasazuki, *Science (New York, N.Y* **260**, 85 (Apr 2, 1993).
5. P. M. Comoglio, L. Trusolino, *The Journal of clinical investigation* **109**, 857 (Apr, 2002).
6. S. Velho *et al.*, *Human molecular genetics* **19**, 697 (Feb 15).
7. D. Brancho *et al.*, *Molecular and cellular biology* **25**, 3670 (May, 2005).
8. S. Arena, A. Pisacane, M. Mazzone, P. M. Comoglio, A. Bardelli, *Proceedings of the National Academy of Sciences of the United States of America* **104**, 11412 (Jul 3, 2007).
9. Y. Samuels *et al.*, *Cancer cell* **7**, 561 (Jun, 2005).
10. P. Michieli *et al.*, *Oncogene* **12**, 775 (Feb 15, 1996).
11. A. Follenzi, L. E. Ailles, S. Bakovic, M. Geuna, L. Naldini, *Nature genetics* **25**, 217 (Jun, 2000).
12. M. Kohli, C. Rago, C. Lengauer, K. W. Kinzler, B. Vogelstein, *Nucleic acids research* **32**, e3 (2004).

Claims

1. An *in vitro* diagnostic method comprising measuring MLK4 gene expression in a tumor cell sample material of a patient, characterized in that measuring MLK4 gene overexpression in said tumor cell sample material is indicative of an invasive potential of said tumor.

2. The *in vitro* diagnostic method according to claim 1, wherein said measuring MLK4 gene expression comprises measuring MLK4 protein expression and/or MLK4 coding nucleic acid, preferably RNA, expression.

3. The *in vitro* diagnostic method according to claim 1 or claim 2, wherein said tumor cell sample material is a bioptic sample.

4. The *in vitro* diagnostic method according to any one of the preceding claims, wherein said bioptic sample corresponds to an invasive portion of said tumor cell sample material.

5. The *in vitro* diagnostic method according to any one of the preceding claims, wherein said measuring MLK4 gene expression comprises measuring wild-type and/or mutated MLK4 gene expression.

6. The *in vitro* diagnostic method according to claim 5, wherein said mutated MLK4 gene contains at least one somatic mutation at the amino acid level selected from H261Y, H261Q, G291E, A293E, W296Stp, R470C, R553Stp, R555Stp, N596I, K629E, P843S, H890P, and M894T mutations.

7. The *in vitro* diagnostic method according to claim 5, wherein said mutated MLK4 gene contains at least one somatic mutation at the nucleic acid level selected from C781T, C783G, G872A, C878A, G888A, C1408T, C1657T, C1663T, A1787T, A1885G, C2788T, A2669C, T2681C mutation.

8. The *in vitro* diagnostic method according to any one of the preceding claims, wherein said measuring MLK4 gene expression is performed using a reagent able to selectively bind to MLK4 protein and/or MLK4 coding nucleic

acid and wherein said reagent is directly or indirectly labeled with a detectable substance.

9. The *in vitro* diagnostic method according to claim 9, wherein said reagent is an antibody specific for MLK4 protein and/or portions thereof or a nucleic acid probe specific for MLK4 nucleic acid and/or fragments thereof.

10. The *in vitro* diagnostic method according to any one of the preceding claims, wherein said method further comprises measuring at least one KRAS gene molecular alteration in said tumor cell sample material, wherein said measuring at least one KRAS gene alteration comprises measuring KRAS protein and/or RNA expression and/or mutations in the nucleotide sequence of the KRAS gene.

12. The *in vitro* diagnostic method according to any one of the preceding claims, wherein said tumor is a colorectal, bladder, breast, gastric, melanoma, lung, ovary or GMB tumor.

13. The *in vitro* diagnostic method according to any one of claims 5 to 12, wherein measuring mutated MLK4 gene overexpression allows to predict whether said patient will be nonresponsive to treatment with EGFR specific binding agents, preferably EGFR specific antibodies.

14. A therapeutic approach involving the reduction or abrogation of MLK4 RNA or protein expression or the use of inhibitors of MLK4 enzymatic activity to treat patients whose tumors bear mutations in the nucleotide sequence of the gene KRAS.

15. Antagonists and/or inhibitors of MLK4 expression and/or MLK4 enzymatic activity for use in the treatment of a tumor, wherein said tumor bears at least one mutation in the KRAS gene encoded protein, wherein said at least one mutation of KRAS gene encoded protein is selected among G12S, G12V, G12D, G12A, G12C, G13A and G13D.

Figure 1

a

Tumor type	Nt change (cDNA)	Aa change (Protein)	Exon	Mutation type	Conserved residue	Ref
CRC	c. C781T	p. H261Y	1	Missense	x	5
CRC	c. C783G	p. H261Q	1	Missense	x	5
CRC	c. G872A	p. G291E	2	Missense	x	5
CRC	c. C878A	p. A293E	2	Missense	x	5
CRC	c. G888A	p. W296Stp	2	Nonsense	x	5
CRC	c. C1408T	p. R470C	5	Missense	x	5
CRC	c. C1408T	p. R470C	5	Missense	x	5
CRC	c. C1657T	p. R553Stp	6	Nonsense	x	5
GBM	c. C1663T	p. R555Stp	6	Nonsense	x	(This study)
CRC	c. A1787T	p. N596I	7	Missense	x	5
CRC	c. A1885G	p. K629E	8	Missense		5
GBM	c. C2788T	p. P843S	9	Missense		(This study)
GBM	c. A2669C	p. H890P	9	Missense		6
GBM	c. T2681C	p. M894T	9	Missense		6

b

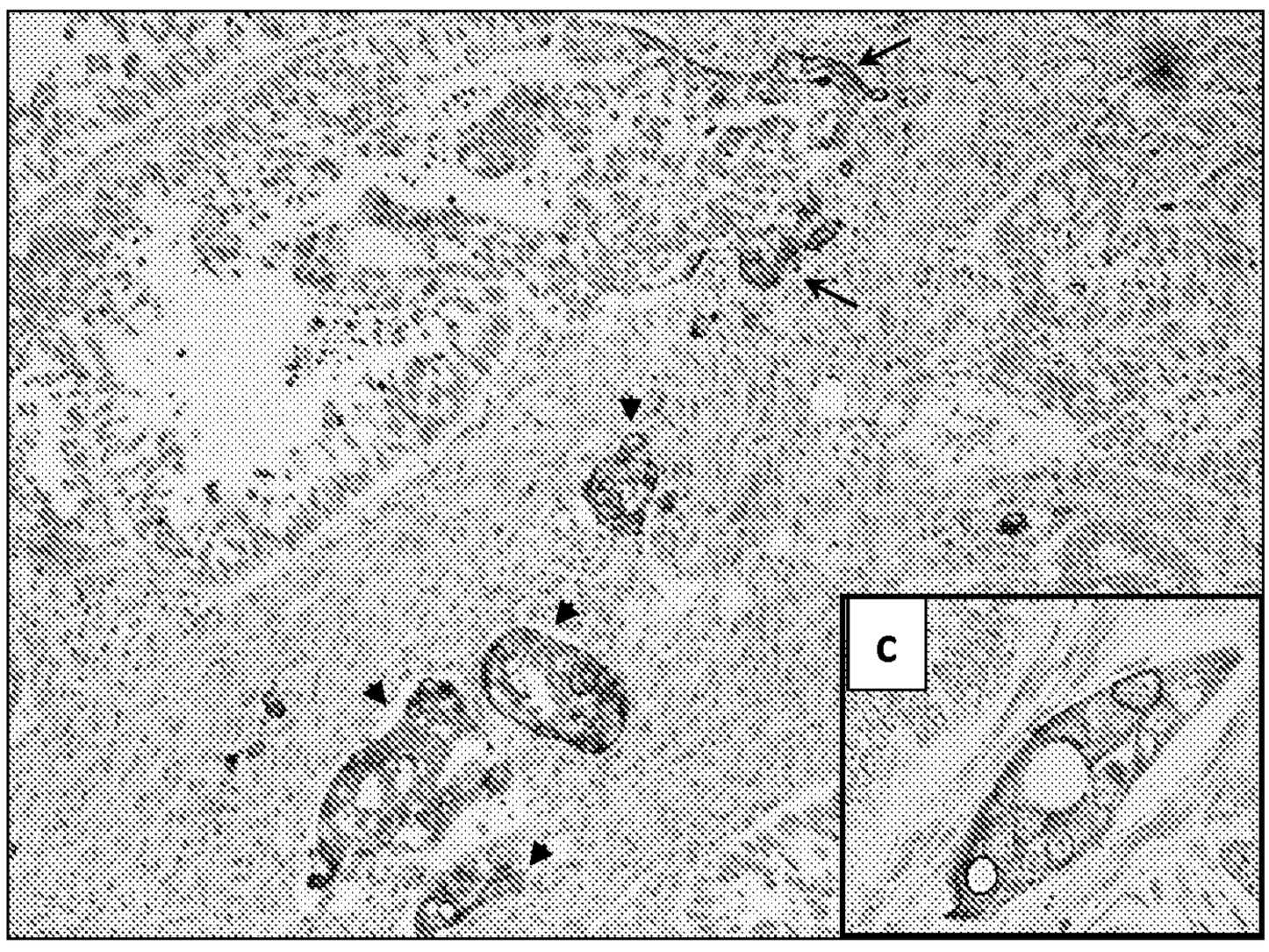


Figure 1/cont.

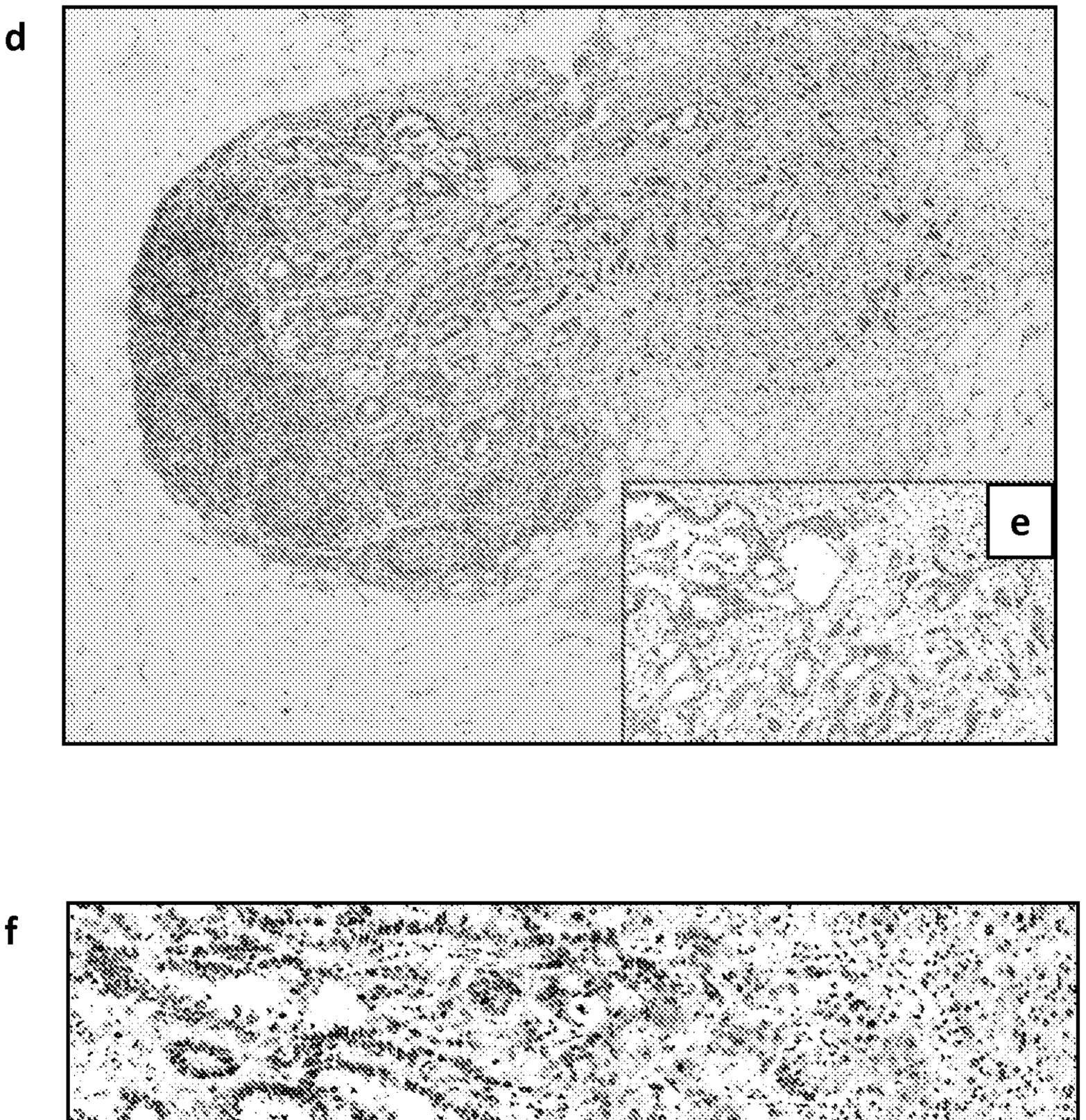
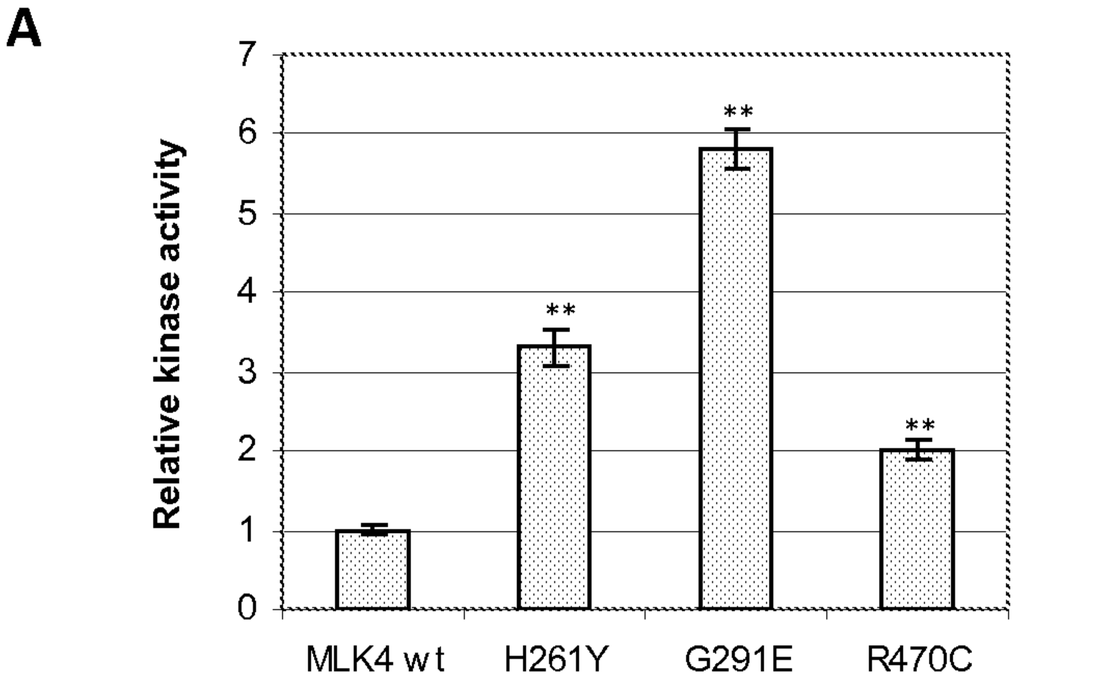
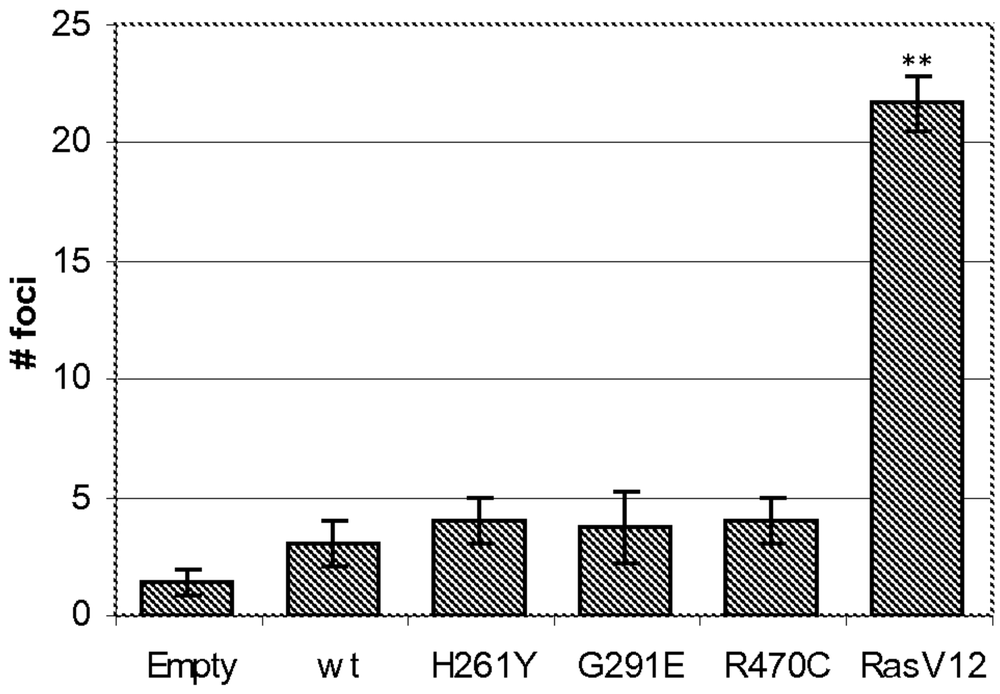


Figure 2


A
Relative kinase activity
0
1
2
3
4
5
6
7
**
**
**
MLK4 wt
H261Y
G291E
R470C


B
foci
0
5
10
15
20
25
**
Empty
wt
H261Y
G291E
R470C
RasV12

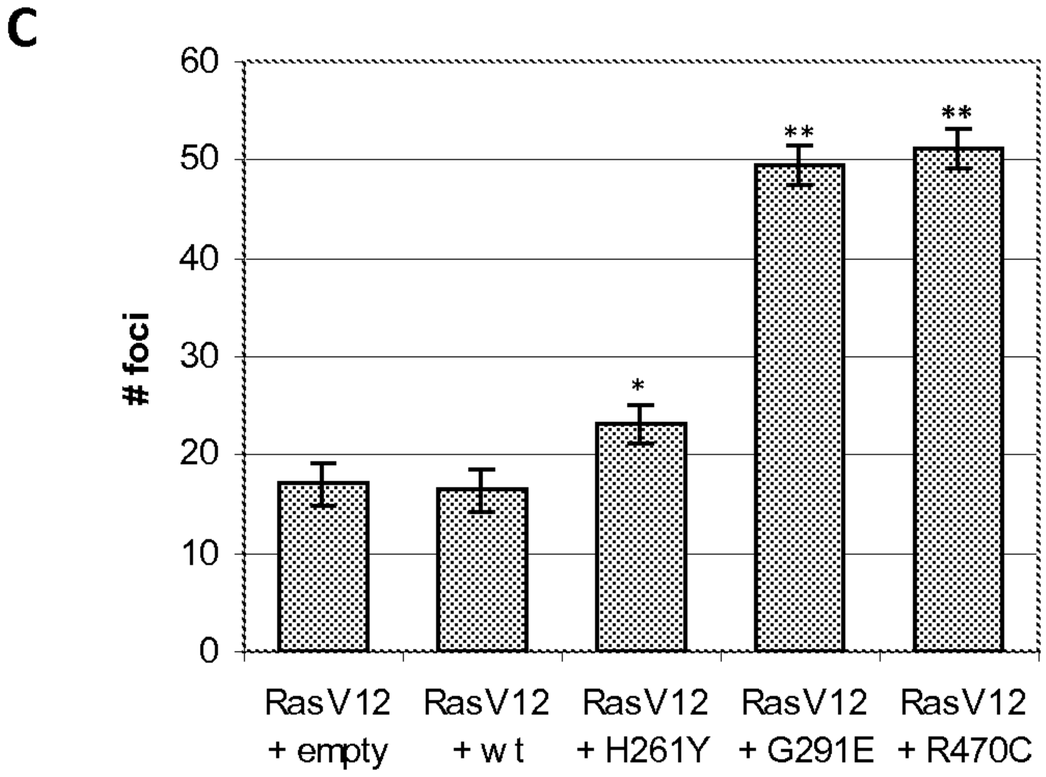
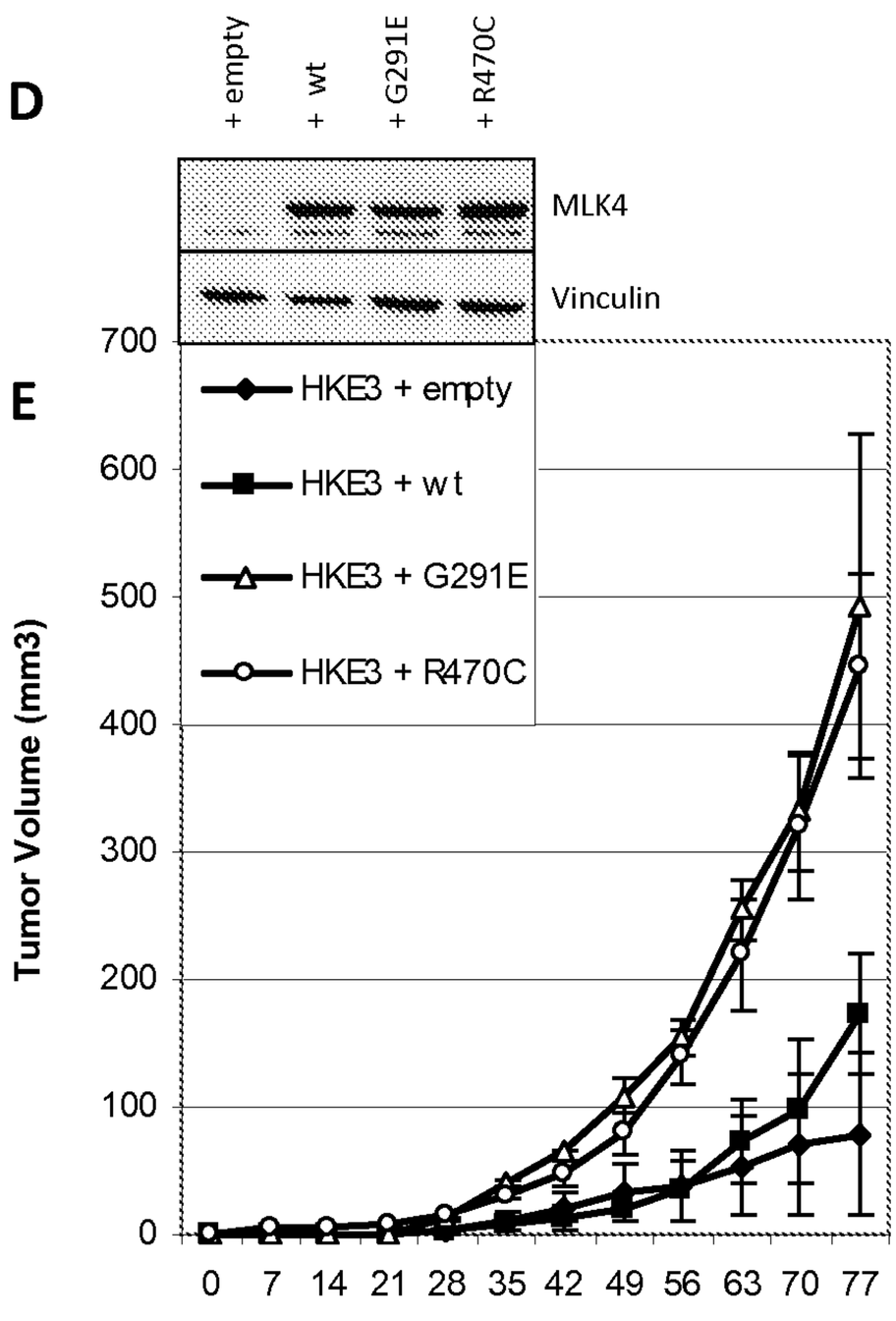

C
foci
0
10
20
30
40
50
60
*
**
**
RasV12 + empty
RasV12 + wt
RasV12 + H261Y
RasV12 + G291E
RasV12 + R470C

Figure 2/cont.



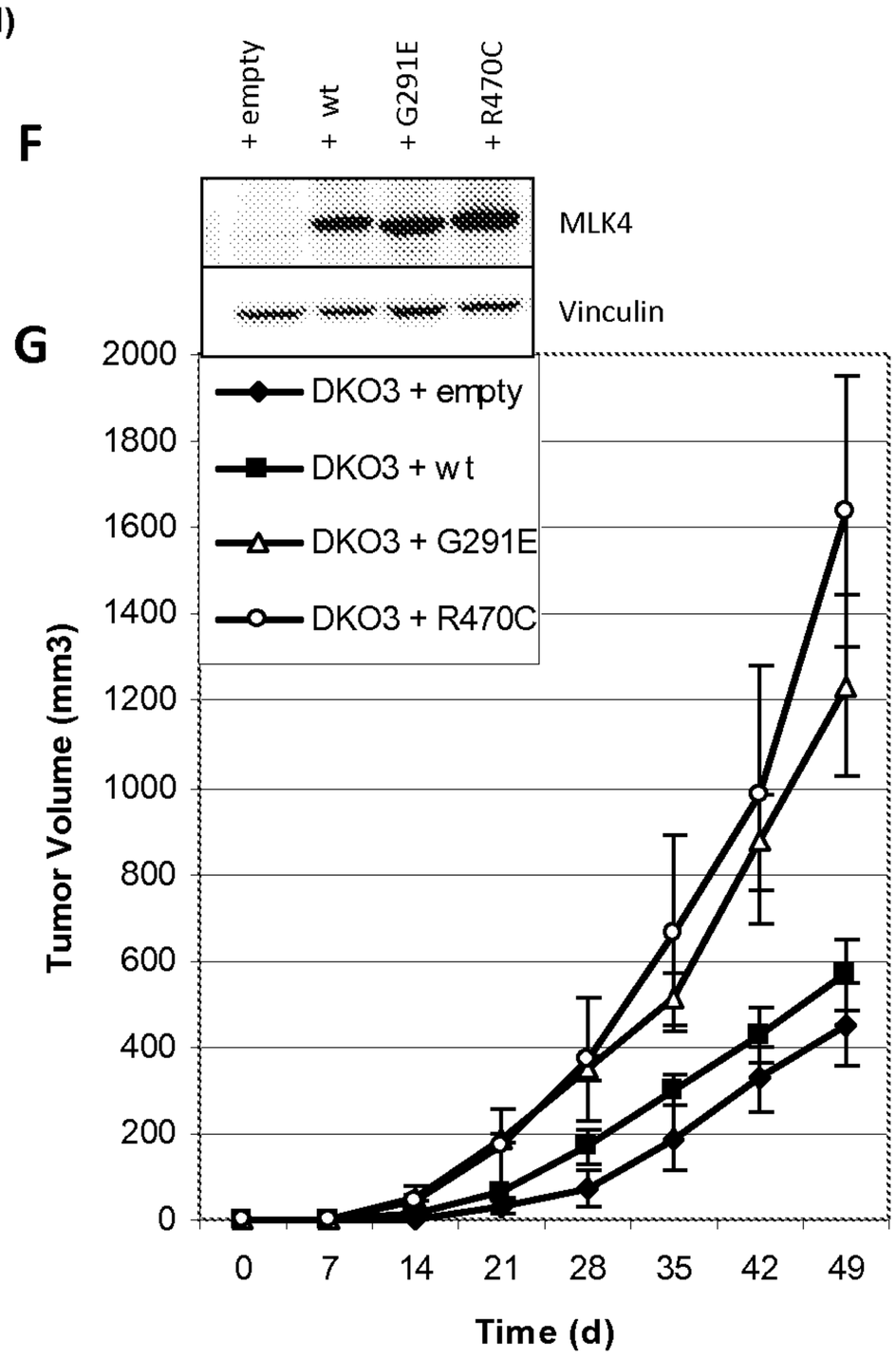
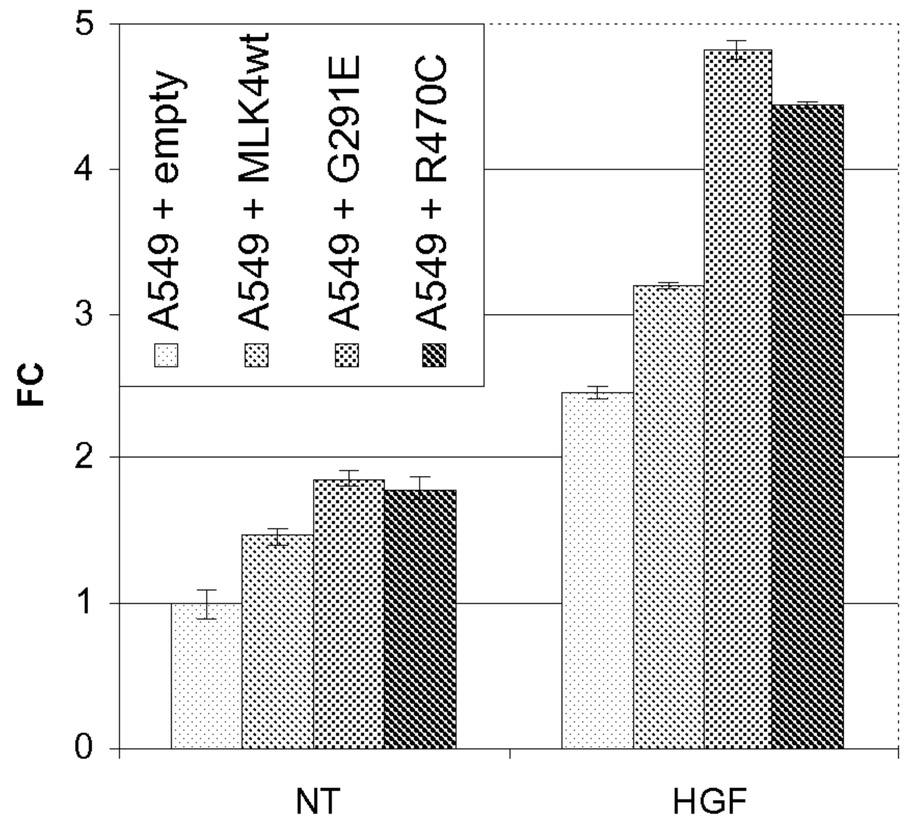
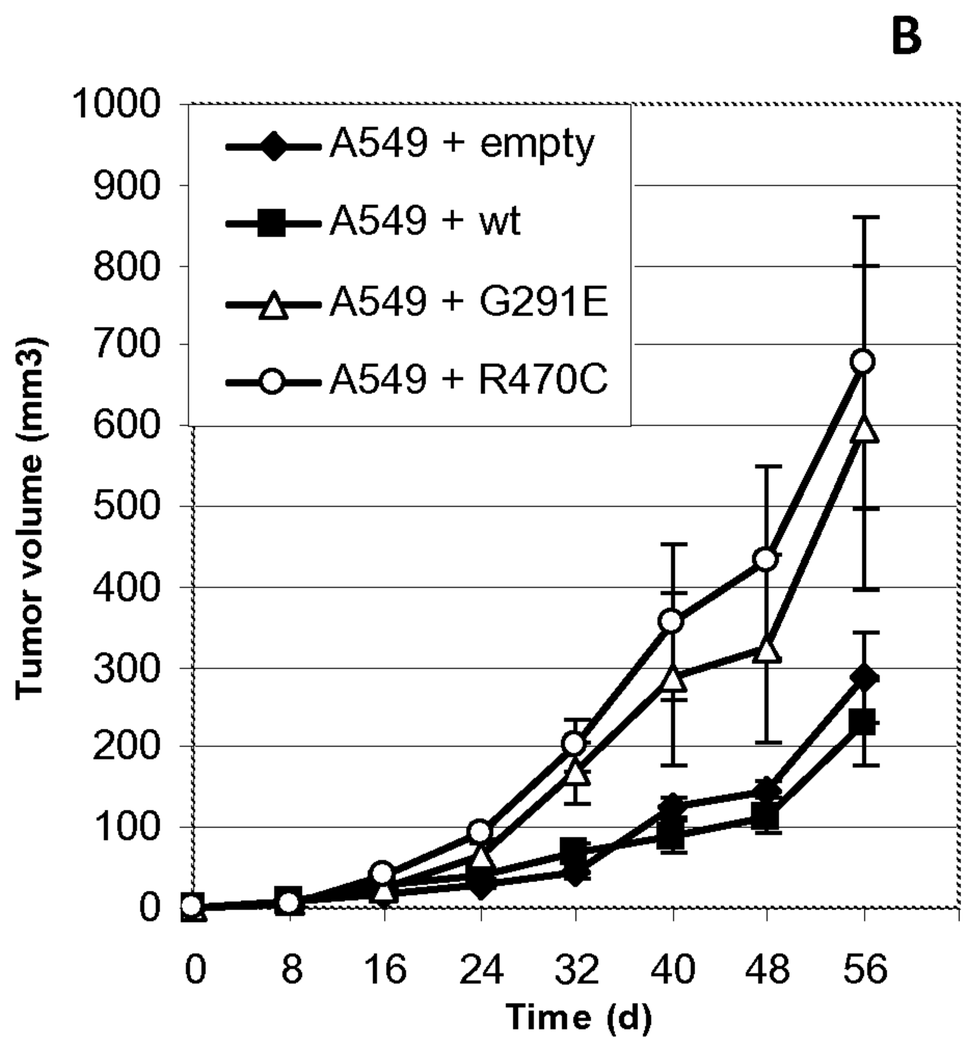
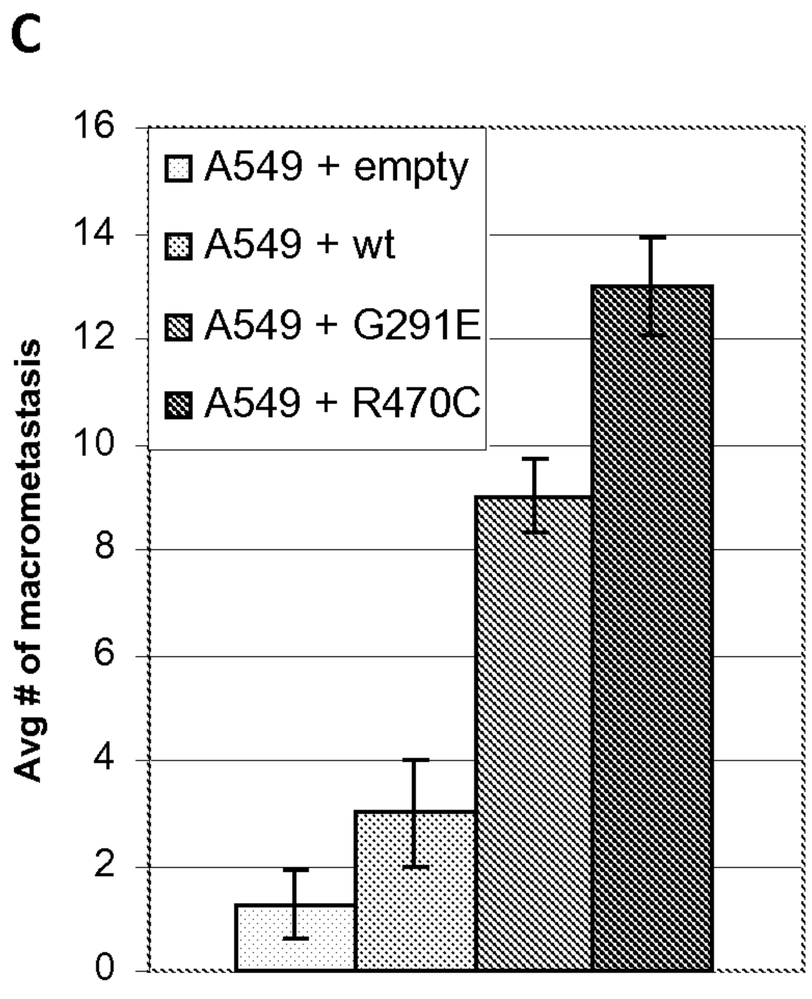


Figure 3


A
FC
A549 + empty
A549 + MLK4wt
A549 + G291E
A549 + R470C
NT
HGF


B
Tumor volume (mm3)
A549 + empty
A549 + wt
A549 + G291E
A549 + R470C
Time (d)


C
Avg # of macrometastasis
A549 + empty
A549 + wt
A549 + G291E
A549 + R470C

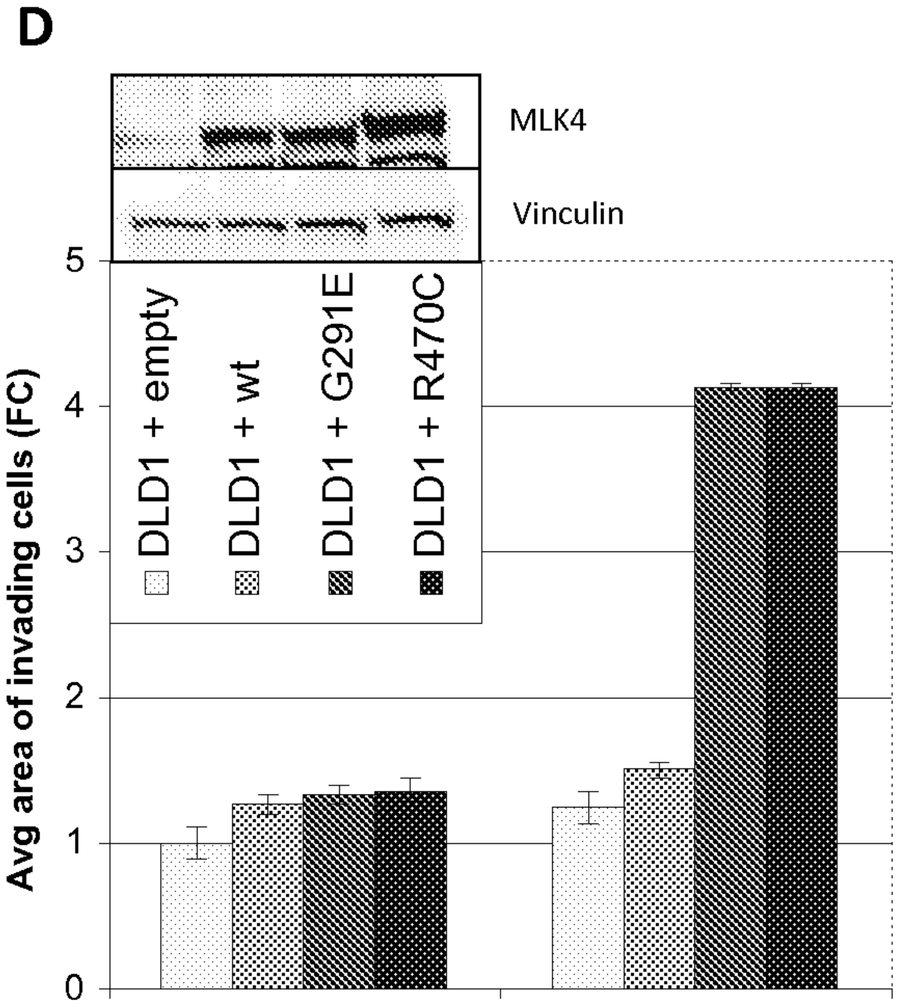
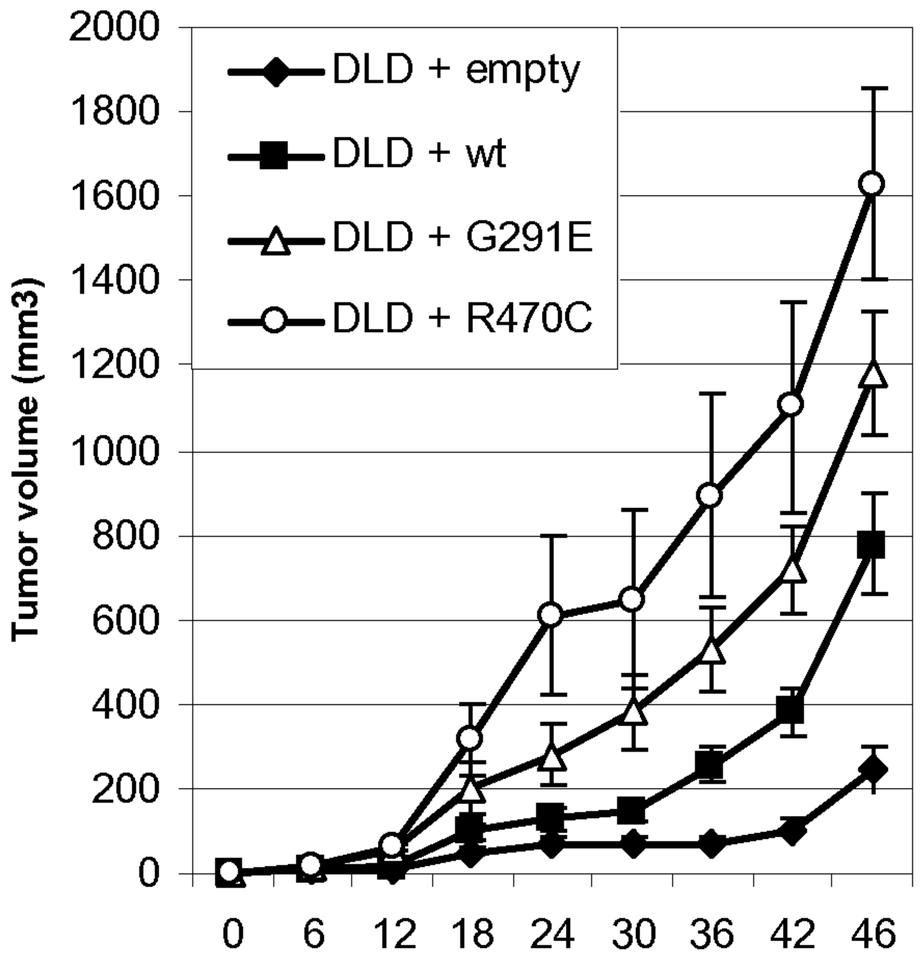

D
MLK4
Vinculin
Avg area of invading cells (FC)
DLD1 + empty
DLD1 + wt
DLD1 + G291E
DLD1 + R470C
NT
HGF

Figure 3/cont.

E



F

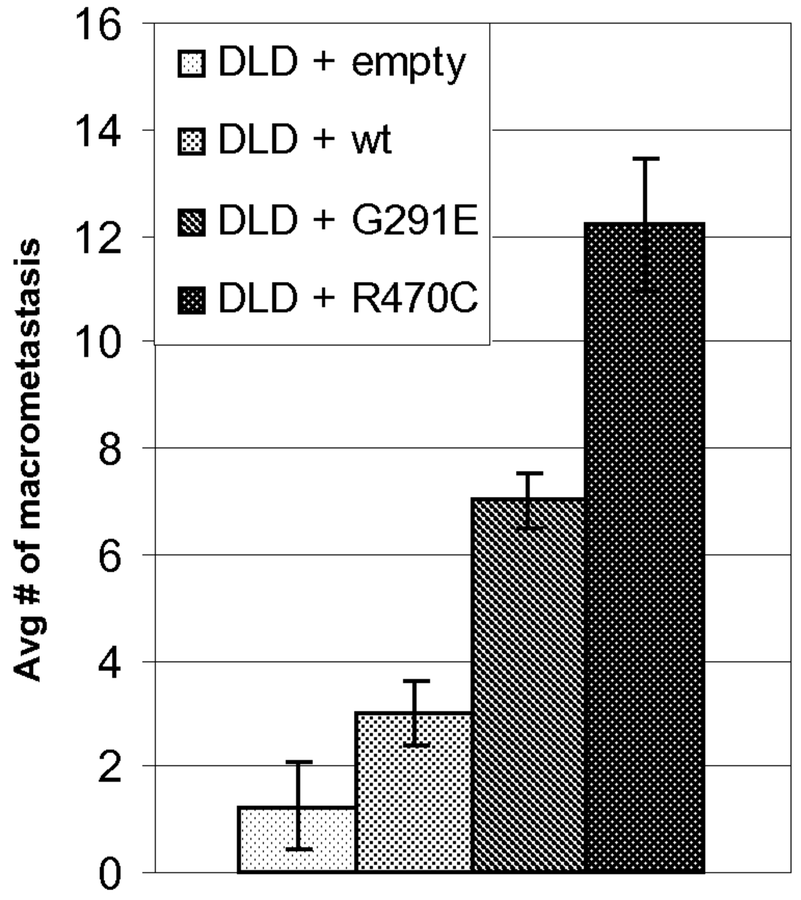


Figure 4

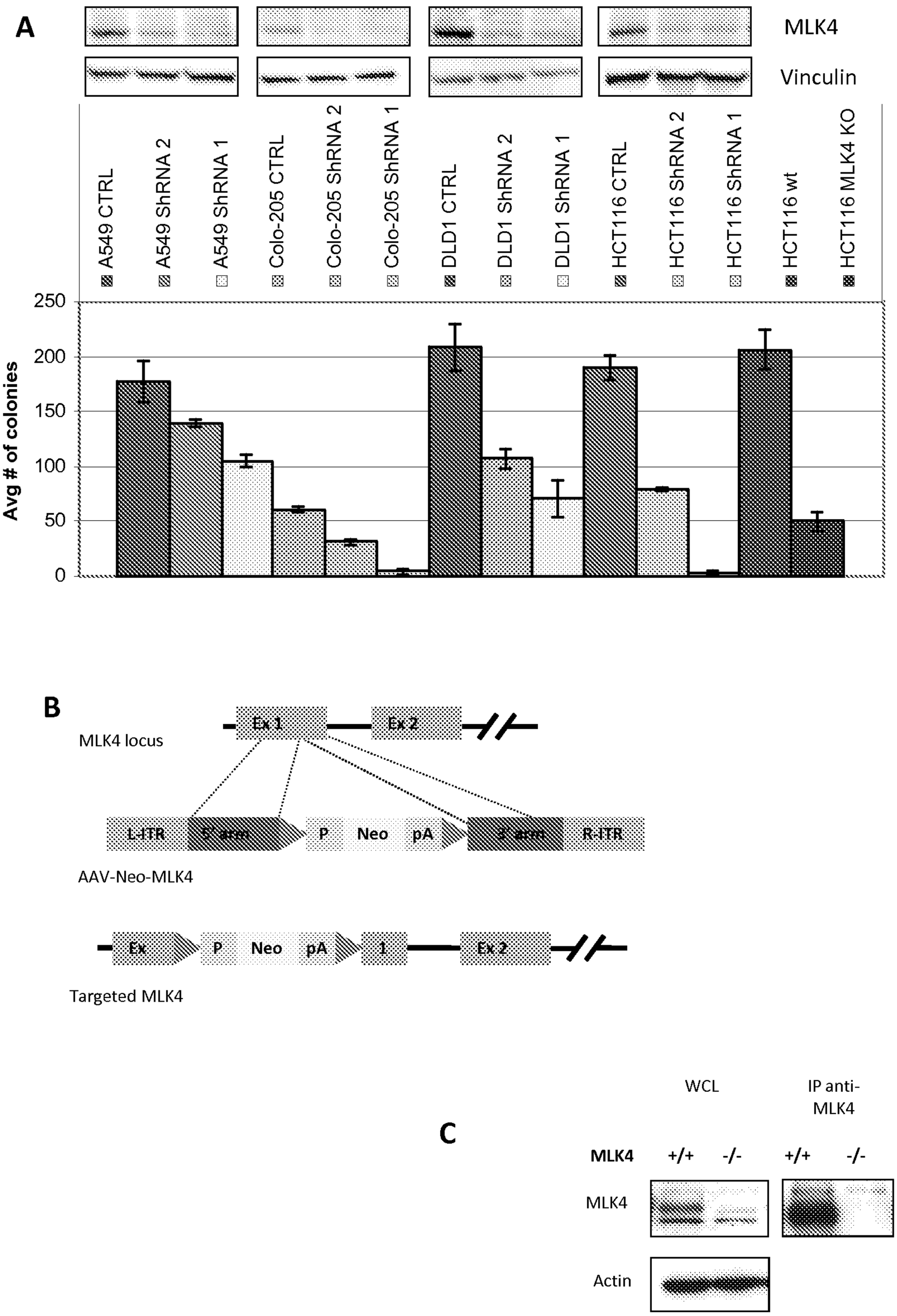
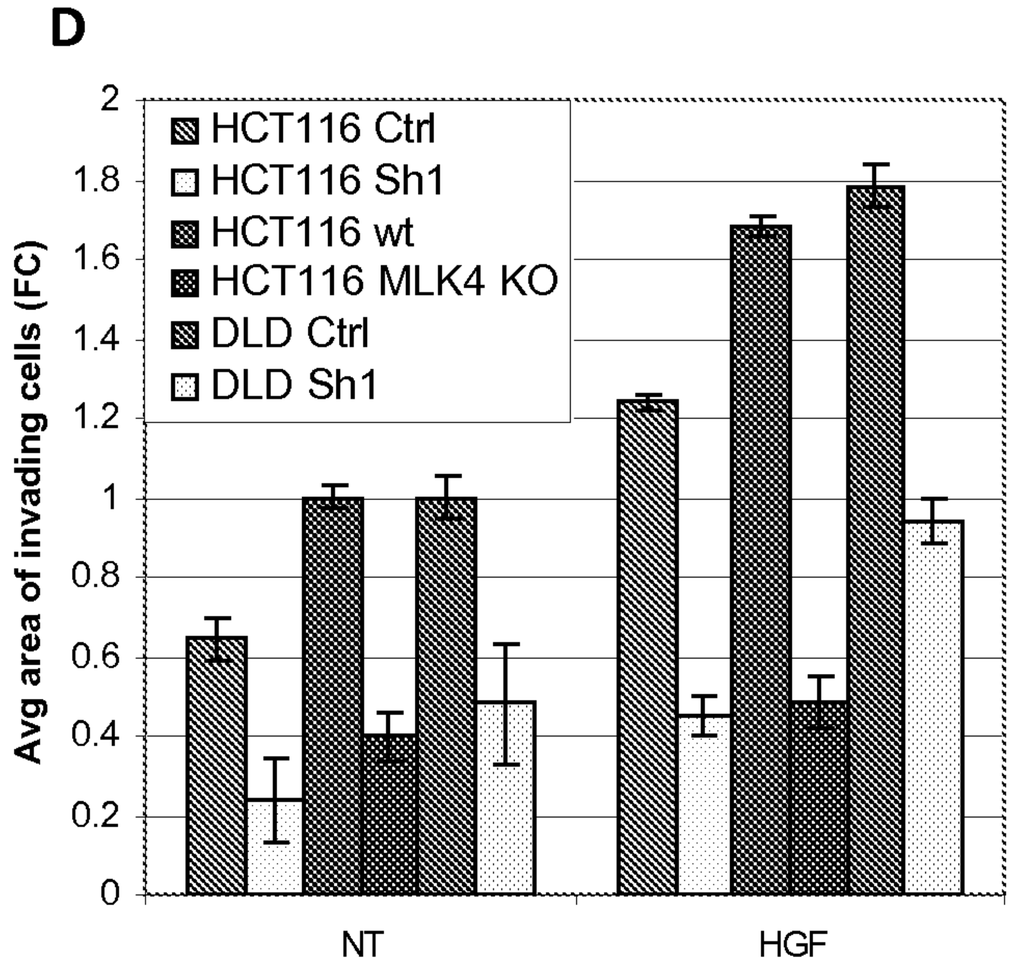
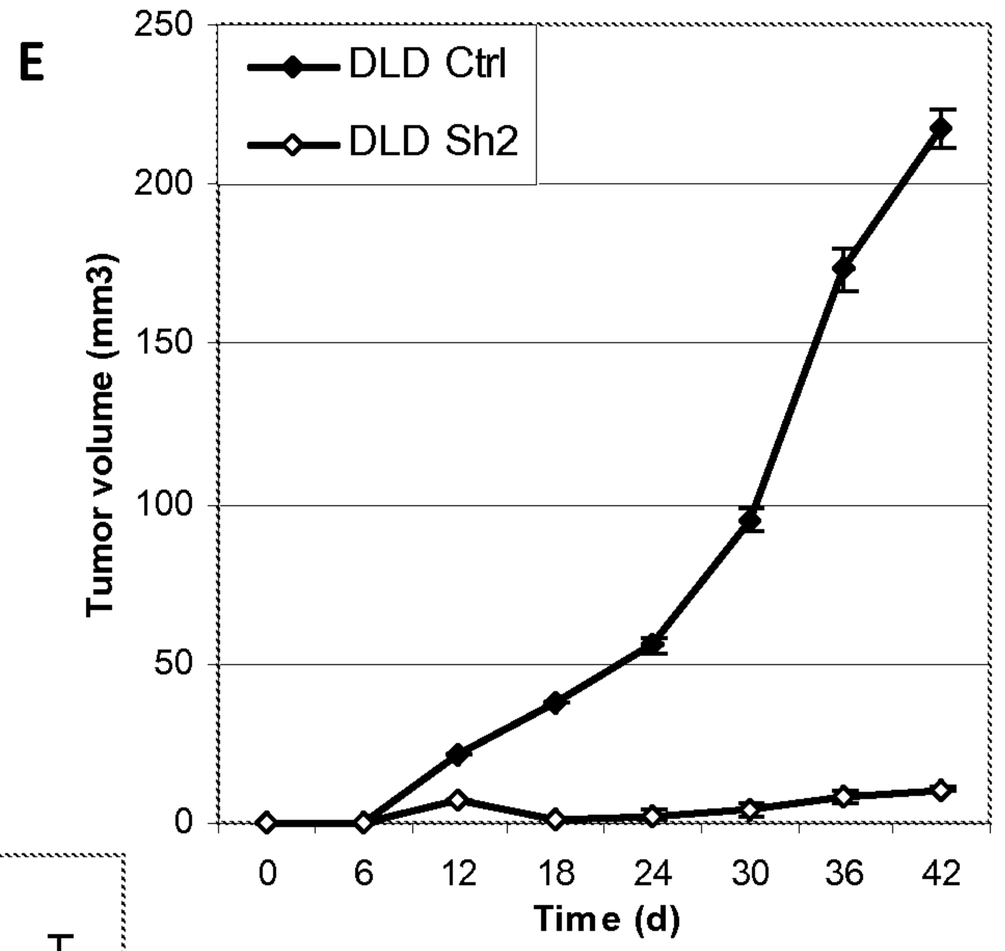

A
MLK4
Vinculin
A549 CTRL
A549 ShRNA 2
A549 ShRNA 1
Colo-205 CTRL
Colo-205 ShRNA 2
Colo-205 ShRNA 1
DLD1 CTRL
DLD1 ShRNA 2
DLD1 ShRNA 1
HCT116 CTRL
HCT116 ShRNA 2
HCT116 ShRNA 1
HCT116 wt
HCT116 MLK4 KO
Avg # of colonies
250
200
150
100
50
0
B
MLK4 locus
Ex 1
Ex 2
L-ITR
5' arm
P
Neo
pA
3' arm
R-ITR
AAV-Neo-MLK4
Ex
P
Neo
pA
1
Ex 2
Targeted MLK4
C
WCL
IP anti-MLK4
MLK4
+/+
-/-
+/+
-/-
MLK4
Actin

Figure 4/cont.





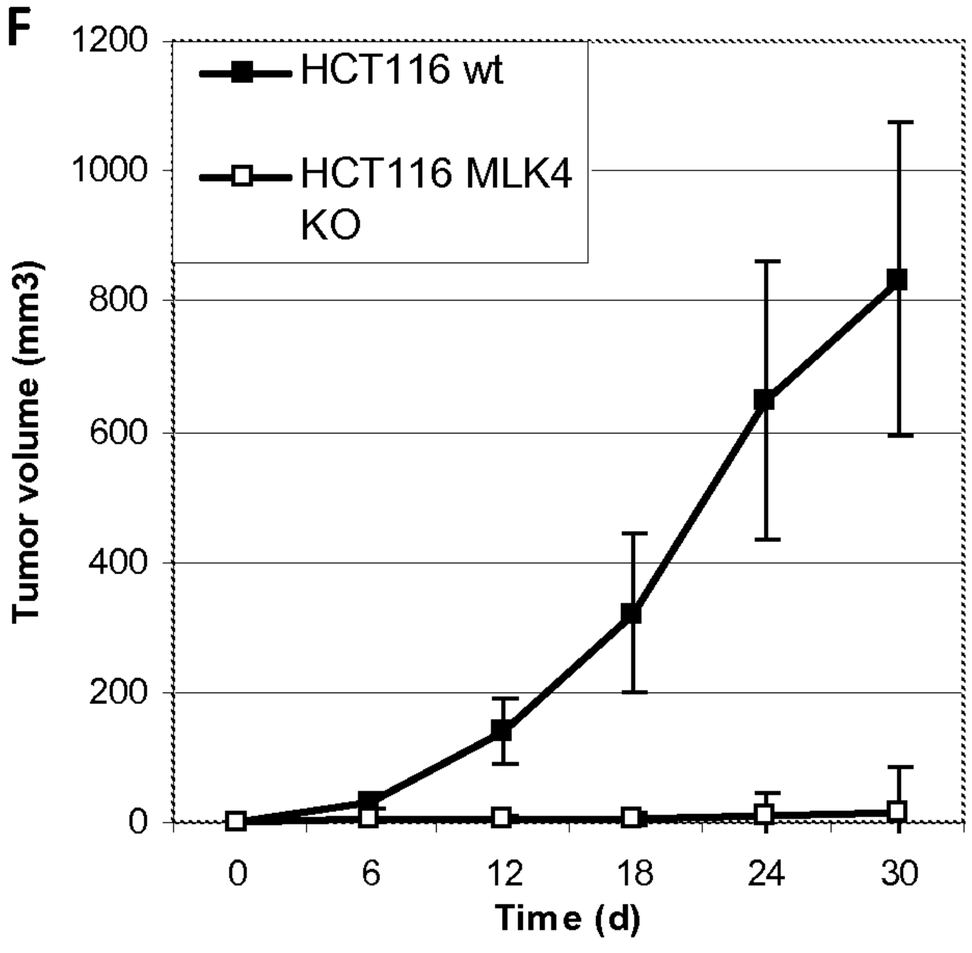


Figure 5

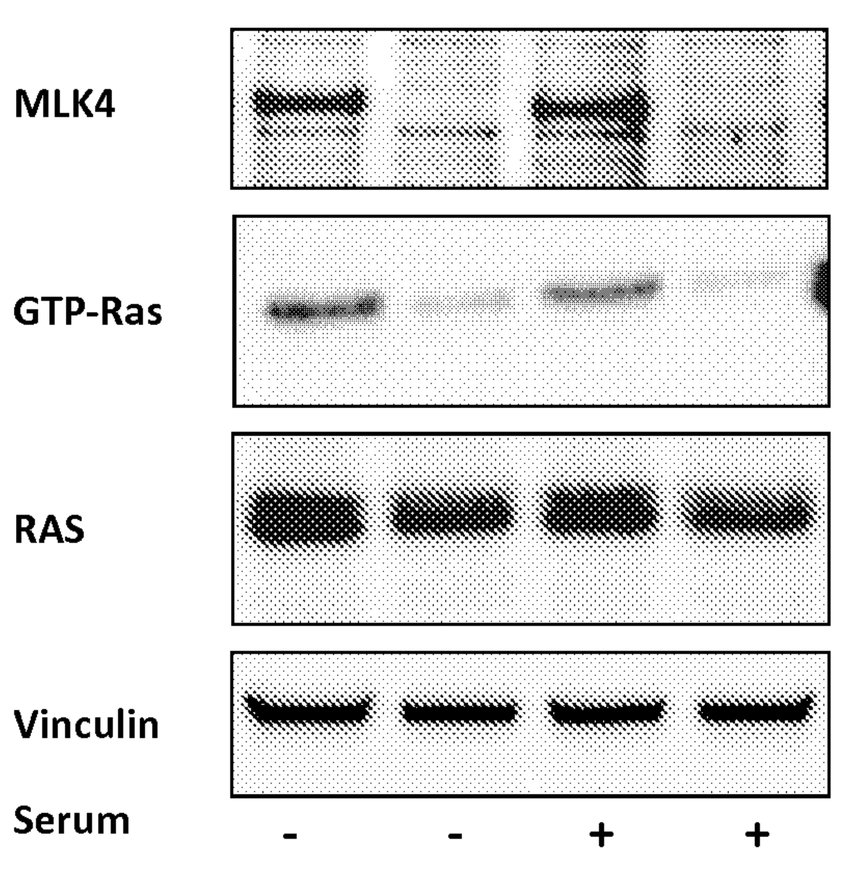
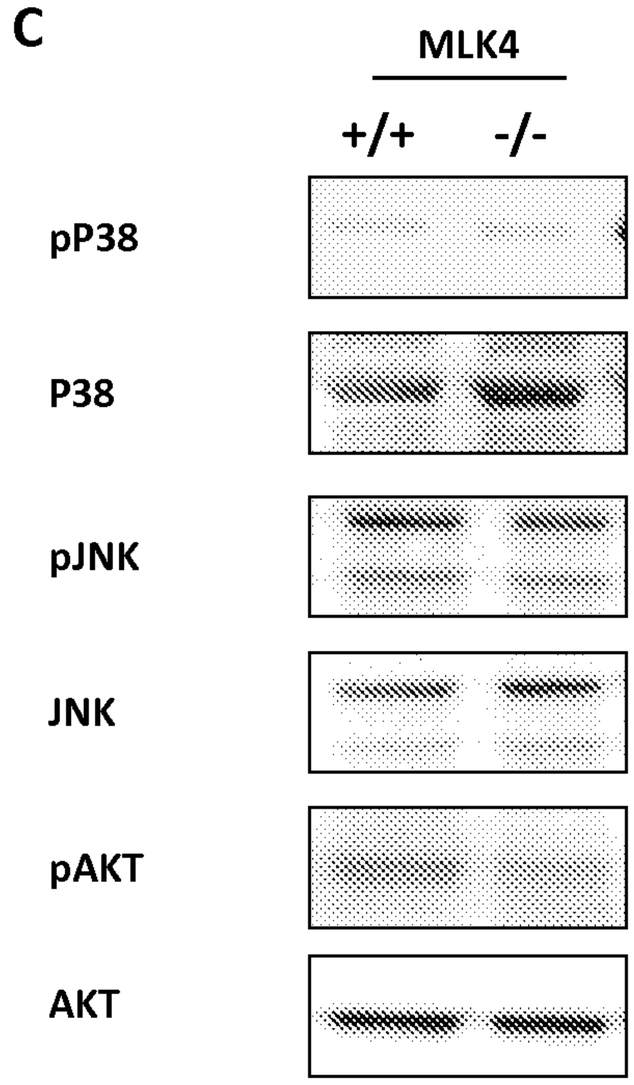
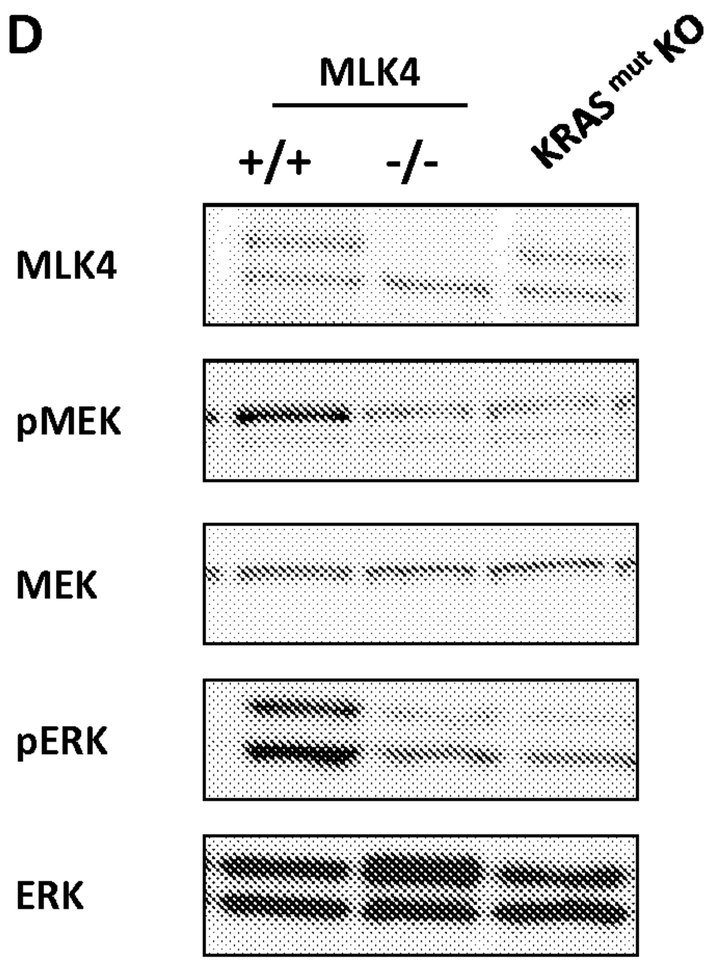

A
MLK4
+/+
-/-
+/+
-/-
MLK4
GTP-Ras
RAS
Vinculin
Serum
-
-
+
+
B
MLK4
+/+
-/-
MLK4
pMEK
MEK
pERK
ERK

Figure 5/cont.





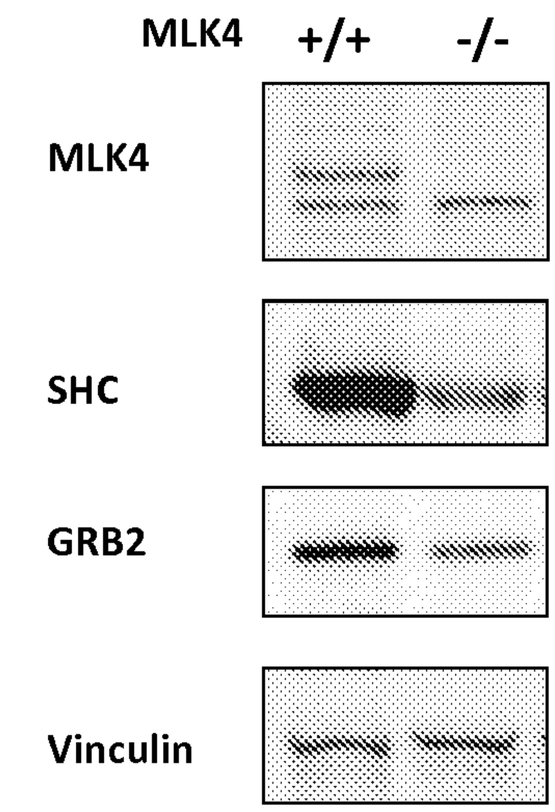
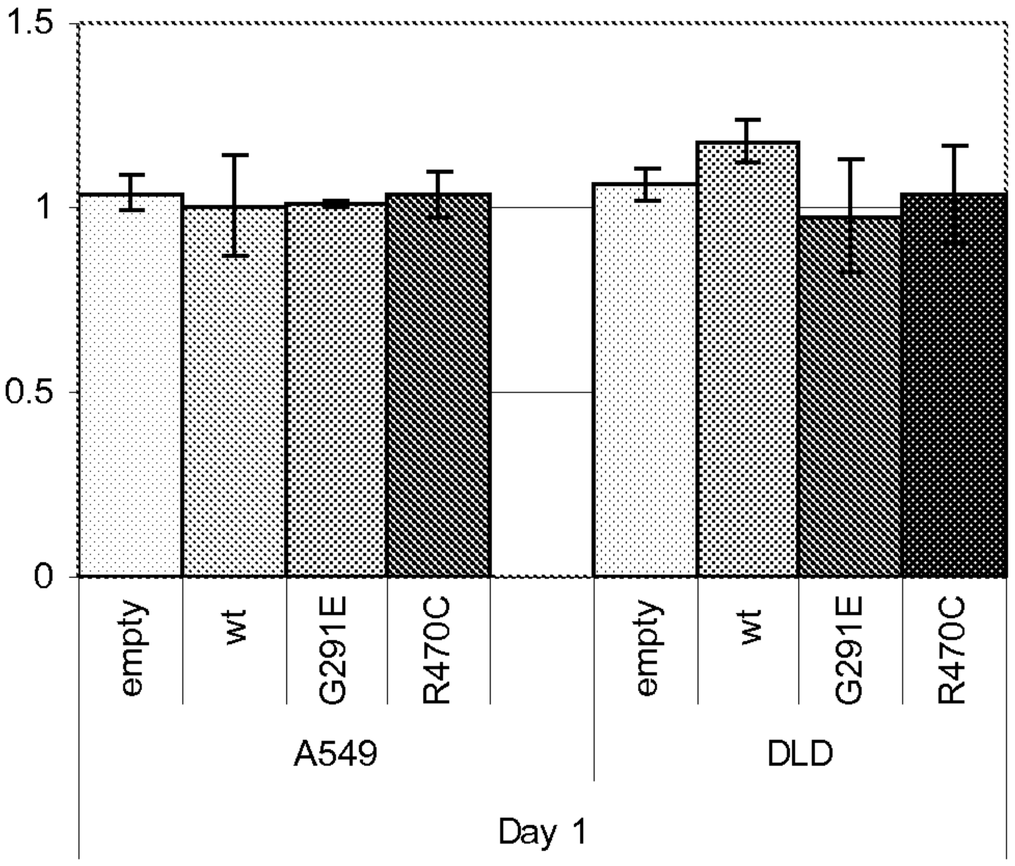
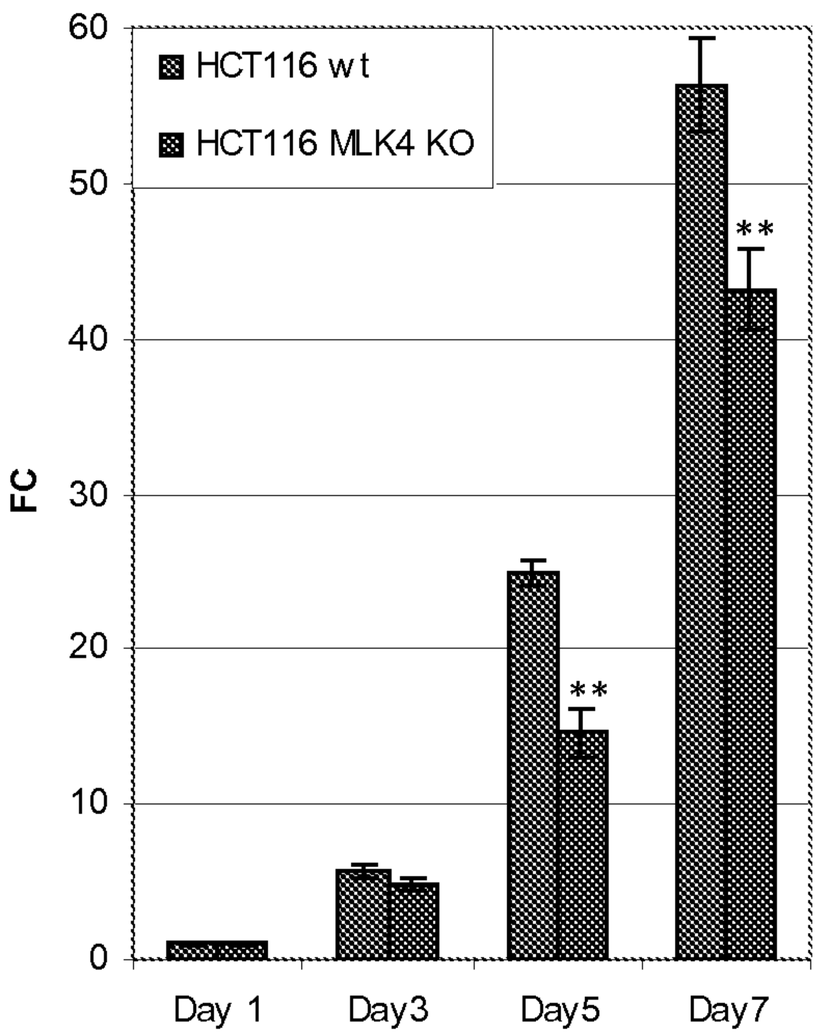


Figure 6

A


1.5
1
0.5
0
FC
empty
wt
G291E
R470C
empty
wt
G291E
R470C
A549
DLD
Day 1

B


60
50
40
30
20
10
0
FC
HCT116 wt
HCT116 MLK4 KO
**
**
Day 1
Day3
Day5
Day7

C

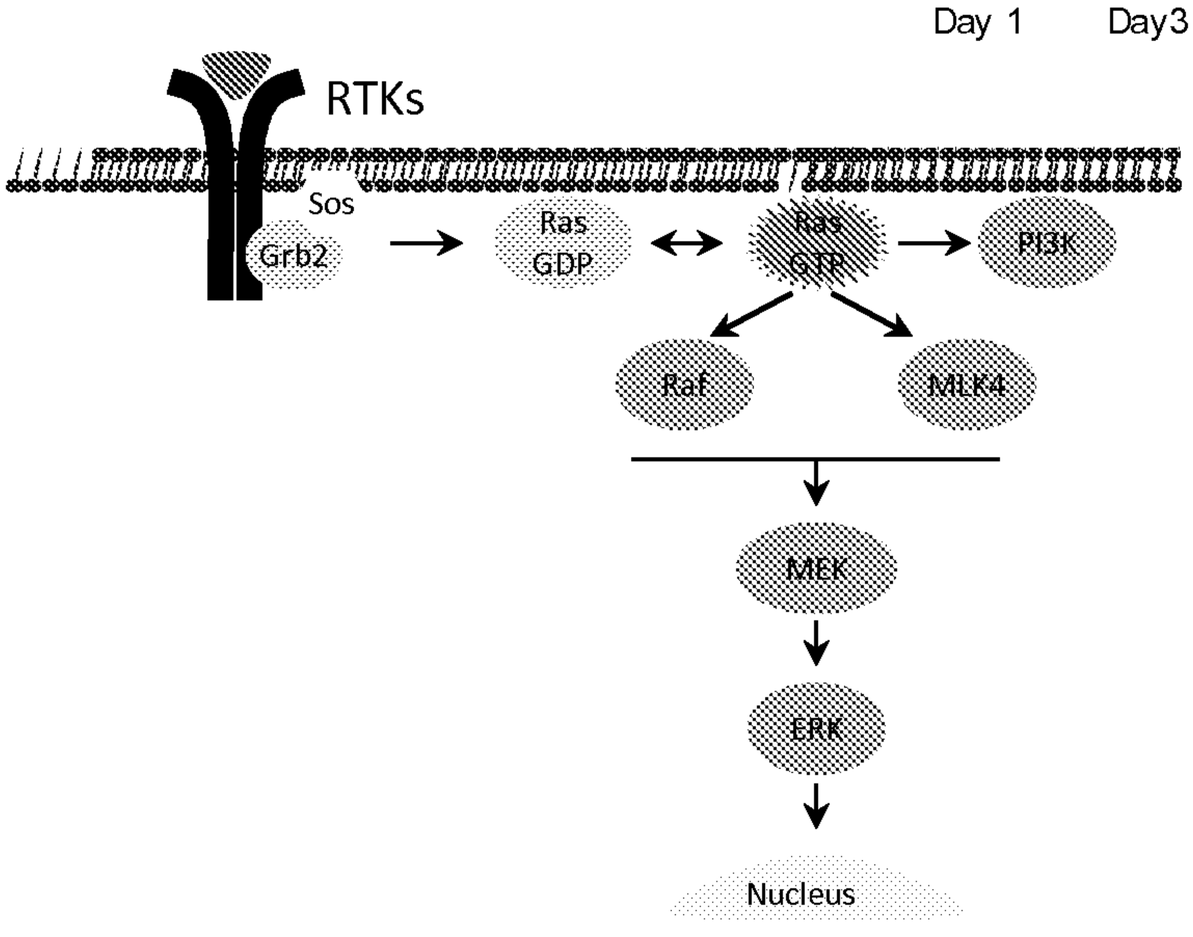
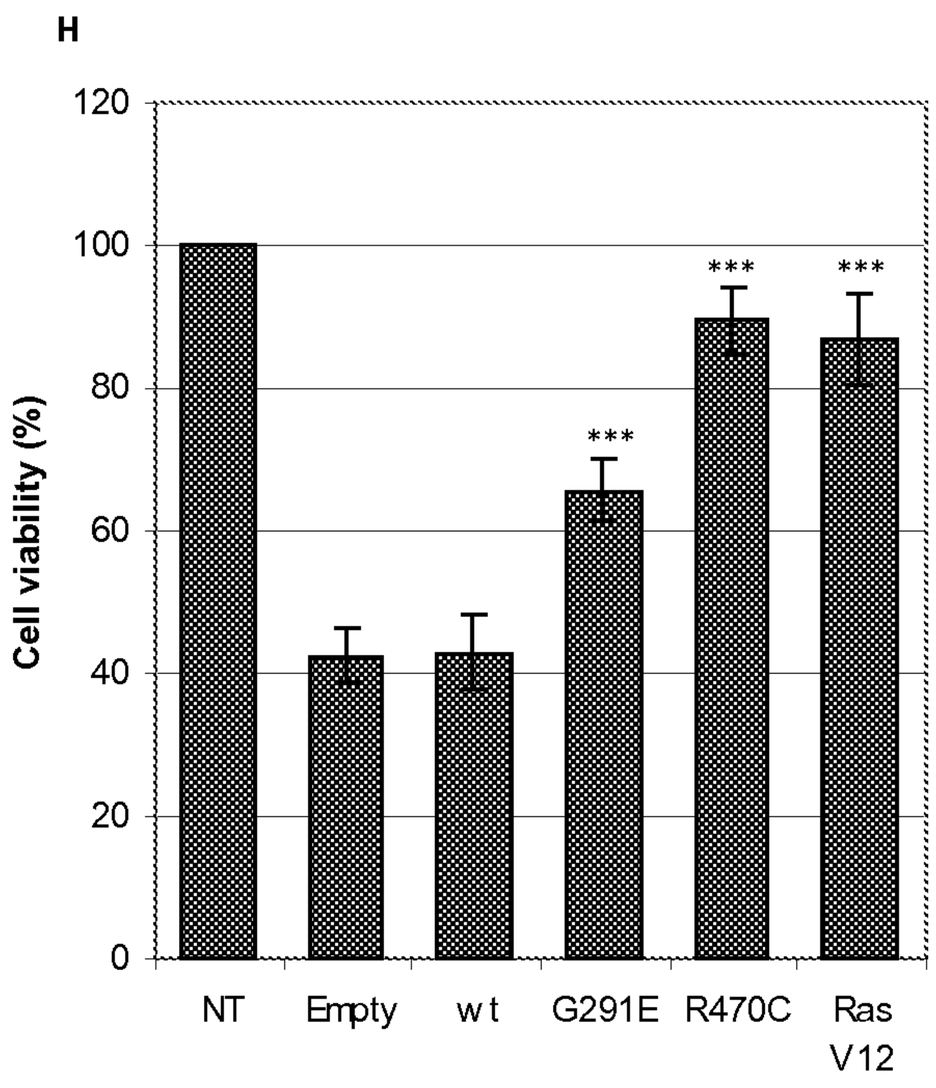

RTKs
Sos
Grb2
Ras
GDP
Ras
GTP
PI3K
Raf
MLK4
MEK
ERK
Nucleus

Figure 7


H
120
100
80
60
40
20
0
Cell viability (%)
***
***
***
NT
Empty
wt
G291E
R470C
Ras V12