Genomic and biological characterization of exon 4 KRAS mutations in human cancer

Manickam Janakiraman[1], Efsevia Vakiani[1,2], Zhaoshi Zeng[3], Christine A. Pratilas[4], Barry S. Taylor[5], Dhananjay Chitale[1,2], Ensar Halilovic[6], Manda Wilson[5], Kety Huberman[1], Julio Cezar Ricarte Filho[1], Yogindra Persaud[1], Douglas A. Levine[3], James A. Fagin[1,7], Suresh C. Jhanwar[2], John M. Mariadason[8], Alex Lash[5], Marc Ladanyi[1,2], Leonard B. Saltz[7], Adriana Heguy[1], Philip B. Paty[3], David B. Solit[1,7*]

[1]Human Oncology and Pathogenesis Program and the Departments of [2]Pathology, [3]Surgery, [4]Pediatrics, [5]Computational Biology, [6]Molecular Pharmacology and Chemistry, and [7]Medicine, Memorial Sloan-Kettering Cancer Center, New York, NY and the [8]Ludwig Institute for Cancer Research, Melbourne, Australia,

Running Title: Exon 4 KRAS mutations in human cancers.

SUPPLEMENTAL METHODS

Cell culture and Immuno blotting. Cancer cell lines were obtained as specified in Supplemental Table 2. LS1034, HCC2998, CCCL-18, and CCCL-23 were grown in MEM with 10mM HEPES, non-essential amino acids (NEAA); C32, C80, C84 and C125PM in Iscove's modified DMEM; and SW1222 and SW1417 were grown using αMEM. All cell lines were supplemented with 2mM glutamine, 50units/ml each of penicillin and streptomycin, and 10% heat inactivated fetal bovine serum (Gemini Bioproducts, Carlsbad, CA) and incubated at 37°C in 5% CO_2. For proliferation assays, cells were plated in 96 well plates, at a density of 2500-5000 cells per well. After 24 hours, cells were treated with PD0325901 or gefitinib at a range of concentrations prepared by serial dilution. PD0325901 and gefitinib for *in vitro* studies was dissolved in DMSO to yield 1 and 10mM stock solutions, respectively, and stored at -20°C. For proliferation studies, the cells were exposed to Alamar Blue (AccuMed International, OH) three to five days following drug treatment, and plates were read using a fluorescence spectrophotometer. The dose required to inhibit growth by 50% (IC_{50}) was calculated using the SoftMaxPro ver.5 software. For focus formation assays, $5x10^5$ cells growing in log phase were either mock transfected with transfection reagent alone or with non-targeting siRNA pool, or KRAS siRNA pool (Dharmacon, CA). Cells were incubated at 37ºC for 2 weeks. Colonies were then stained with crystal violet (Sigma-Aldrich, St. Louis, MO) for 1 hour and plates were scanned using flat-bed scanner with foci counted manually and graphed.

Immunblots were performed as previously reported (19). p42/44 MAPK, phospho-p42/44 MAPK, Akt, phospho-Akt (ser473), RB, cleaved PARP, and cleaved caspase-3 antibodies were

obtained from Cell Signaling Technology. Anti-Cyclin D1 and p27 antibodies were obtained from Santa Cruz Biotechnology. To measure apoptotic cells, cells were seeded in 10cm dishes at a density of 1 x 10^6 cells/dish and the following day were treated with the indicated concentration of drug or vehicle (DMSO) for the indicated times. Both adherent and floating cells were harvested and stained with ethidium bromide using the method of Nüsse [48].

Additional Sequenom method details. The Sequenom MassARRAY system is based on matrix-assisted laser desorption/ionization-time of flight mass spectrometry (MALDI-TOF MS). In these assays, the mutant and germline alleles for a given point mutation produce single-allele base extension reaction products of different masses that are then resolved by MALDI-TOF MS. Both the amplification and extension primers were designed using Sequenom Assay Designer v3.1 software (Sequenom, San Diego, CA). The amplification primers were designed with a 10mer tag sequence to increase their mass so that they fall outside the range of detection of the MALDI-TOF mass spectrometer. Results were generated using the SpectroTYPER v3.4 software (Sequenom, San Diego, CA). All the positive cases were confirmed by visually reviewing the spectra. For the PCR amplification, a total of 15 ng of genomic DNA (in 1 μl) was amplified in a 5 μl reaction mixture containing 0.1 μl (0.5 U) HotStarTaq enzyme (Qiagen, Valencia, CA), 0.625 μl of 10x HotStar buffer, 0.325 μl of 25 mM (total) MgCl2, 0.25 μl of 10mM (each) deoxynucleotide triphosphate, 1 μl of 100 nM of each forward and reverse primers and 1.7 μl of water. The PCR step was initiated with a 95°C soak for 15 min, followed by 45 cycles, consisting of 95°C for 20 sec, 56°C for 30 sec, 72°C for 60 sec, and a final extension of 3 min at 72°C. After PCR, the remaining unincorporated dNTPs were dephosphorylated by adding 2 μl of the SAP cocktail, containing 1.33 μl of water, 0.17 μl of reaction buffer (Sequenom, San

Diego, CA) and 0.5 µl of SAP (Sequenom, San Diego, CA). The 384-well plate was then sealed and placed in a thermal cycler with the following conditions: 37°C for 40 min, 85°C for 5 min and then held at 4°C indefinitely. After the SAP treatment, a 2 µl cocktail, consisting of 0.755 µl water; 0.2 µl iPlex 10x buffer (Sequenom, San Diego, CA), 0.2 µl iPlex terminator mix (Sequenom, San Diego, CA); 0.804 µl of 7 µM/ 14 µM (depending on the low vs. high mass primers) extension primer mixture and 0.041 µl iPlex enzyme (Sequenom, San Diego, CA) was added. After the iPlex cocktail addition, the plate was again sealed and placed in a thermal cycler with the following program: 94°C for 2 min followed by 40 cycles of 94°C for 5 sec, [5 cycles (52°C for 5 sec, 80°C for 5 sec) and 72°C for 5 sec]. The reaction mixture was then desalted by adding 16 µl of water and 6 mg cationic resin mixture, SpectroCLEAN (Sequenom, San Diego, CA). The plate was then sealed and placed in a rotating shaker for 20 min to desalt the iPlex solution. Completed genotyping reactions were spotted in nanoliter volumes onto a matrix arrayed silicon chip with 384 elements (Sequenom SpectroCHIP) using the MassARRAY Nanodispenser. SpectroCHIPs were analyzed using the Bruker Autoflex MALDI-TOF mass spectrometer and the spectra were processed using the SpectroTYPER v3.4 software (Sequenom, San Diego, CA). Mutations included in the sequenom assay panel are listed in Table S5.

Multiplexed PCR and MS-based genetic fingerprinting assay and analysis method useful for sample identification, matching and tracking.

1) Selection of SNP assay loci. We downloaded all HapMap [49] non- redundant individual genotype data (organized by chromosome and population) from the HapMap FTP site [50]. After parsing through all loci on all chromosomes and in all HapMap populations, we filtering out any polymorphic loci of length greater than one or containing an 'N' character, flagged bi-allelic

SNP haplotypes appearing in AB:AA:BB ratios of greater than 0.45:0.225:0.225 (i.e., close to Hardy-Weinberg equilibrium for bi-allelic loci, where alleles are found in roughly equal proportions). We then counted the number of populations per locus passing this filter, and sorted the results. The loci with the best counts in each chromosome were retained. Allowing for ties, this resulted in 89 HapMap equally bi-allelic loci in a number of populations, which were also scattered across the genome and from which we could design a multiplexed SNP assay (see Table S6).

2) Multiplexed assay primer design. Forty-two out of the selected 89 SNPs, covering all chromosomes (most chromosomes are covered by 2 SNPs, see Table S6), that were amenable to assay design for mass spectrometry-based genotyping, were selected to design 4 multiplexed assays using the Sequenom Assay Design V.3.1 software (Sequenom, San Diego, CA). Two of the multiplexed wells also contained a sex determination assay based on the amelogenin gene which has two versions, one on the X chromosome and one in the Y chromosome and has areas of divergent sequence between the two versions of the gene [51]. Table S7 lists the selected SNP, specific primers and multiplexing scheme.

3) Sequenom assay. See above additional Sequenom method details

4) Sample matching and clustering tool. We created a tool that reads in the assay results and does a pair-wise comparison between all samples to look for samples that might be related. For each sample pair we assume no relation and ask, what is the probability of the two samples having at least *m* matching loci by chance? Since there are only two possible outcomes for two

samples at a locus: match and not matched, and the loci are independent, we used the binomial distribution to calculate the probability of at least *m* matching loci of the loci examined:

$$P = \sum_{k=m}^{N} \binom{N}{k} p^k (1-p)^{N-k}$$

, where *m* is the number of loci that match between the samples, *N* is the total number of loci for which we had results in both samples, and p^k is the probability of a match at a locus. p^k, the probability of two unrelated samples matching at a locus is given by: $(GG \cap GG) \cup (GT \cap GT) \cup (TT \cap TT) = (0.25 \bullet 0.25) + (0.5 \bullet 0.5) + (0.25 \bullet 0.25) = 0.0625 + 0.25 + 0.0625 = 0.375$. If a sample was run more than once and the tool finds more than one sample name and assay identifier pair with the same genotype, the sample name and assay pair is only included once. If the duplicates conflict, all results for that sample and assay pair are excluded. We used the Bonferroni method to correct for multiple comparisons. Any sample pair with a corrected p-value of < 0.05, we consider unlikely to be unrelated. It is also possible that two similar samples have been missed. Sample pairs deemed *likely* to be related are examined manually to determine if a labeling error occurred.

A simple clustering tool reads the results and groups samples that are likely to be related. In this clustering procedure, samples are clustered together in a cluster if they are related (at a corrected p-value of 0.05) to at least one other sample in the cluster. Samples are assigned to one cluster exclusively, and singlet clusters are allowed. Both macros can be downloaded from the web site[52]

5) **Validation of the MS-based genetic fingerprinting assay and analysis method:** In order to cross-validate our assay and compare with other methodologies, we tested DNA from 16 cell

lines that had been previously genotyped by SNP microarray analysis [53]. Seven of the SNPs we used in our assay were represented in the Affymetrix 250K array and only 109 out of the 112 overlapping calls generated on 16 unrelated samples were concordant (p-value = 9.90×10^{-17}, one-tailed Fisher exact test, see Table S8). Additionally, when a panel of colorectal cell lines was analyzed using this assay we found a list of colorectal cancer cell lines that share identical genetic background (see Table S9).

SUPPLEMENTAL REFERENCES

48. Nusse M, Beisker W, Hoffmann C, Tarnok A. Flow cytometric analysis of G1- and G2/M-phase subpopulations in mammalian cell nuclei using side scatter and DNA content measurements. Cytometry 1990; 11: 813-21.
49. The International HapMap Project. Nature 2003; 426: 789-96.
50. http://ftp.hapmap.org/genotypes/2008-11_phaseII+III/forward/non-redundant/
51. Salido EC, Yen PH, Koprivnikar K, Yu LC, Shapiro LJ. The human enamel protein gene amelogenin is expressed from both the X and the Y chromosomes. Am J Hum Genet 1992; 50: 303-16.
52. http://cbio.mskcc.org/dna_fingerprinting/
53. Schweppe RE, Klopper JP, Korch C, et al. Deoxyribonucleic acid profiling analysis of 40 human thyroid cancer cell lines reveals cross-contamination resulting in cell line redundancy and misidentification. J Clin Endocrinol Metab 2008; 93: 4331-41.

SUPPLEMENTAL FIGURE LEGENDS

Supplemental Figure 1. Representative Sanger traces of tumor and germline DNA confirming that A146T and K117N mutations were somatic. Both normal and tumor DNA were sequenced to confirm the somatic nature of the exon 4 mutations. Representative Sanger sequencing traces from normal and tumor DNA are shown for codon A146 (**A)** and K117N (**B**). **C.** Cell lines for which matched tumor tissue was available were shown to harbor identical RAS mutations. Representative Sanger sequencing traces (left) and mass spectrometry peaks (right) are shown for the cell line CCCL-23 (above) and tumor (below).

Supplemental Figure 2. Kaplan-Meier plots of disease-specific survival of 186 patients with stages 1-3 colorectal cancer as a function of KRAS/NRAS mutational status. **A.** Patients were divided into three cohorts: (1) those with KRAS G12/G13 mutations (shown in blue), (2) those with non-G12/G13 mutations in KRAS or in NRAS (shown in red) and (3) those wild-type for KRAS (shown in green). Patients with exon 3 and 4 KRAS or NRAS mutations had a statistically significant improvement in disease-specific survival (p-value = 0.006). A trend towards improved disease-specific survival was observed in a pairwise comparison of the KRAS wild-type cohort versus the cohort of patients with G12/G13 KRAS mutations but this did not reach statistical significance (p-value = 0.07). A statistically significant improvement in disease-specific survival was observed in a pairwise comparison of the G12/G13 KRAS mutant cohort versus the cohort of patients with non-G12/G13 KRAS mutations or NRAS mutations (p-value = 0.02). All differences in disease-specific survival among any two classes of KRAS mutation were tested by the log-rank test statistic. **B.** Patients were divided into two cohorts: (A) KRAS G12/G13 mutant (shown in blue) or (B) KRAS G12/G13 wild-type (shown in green).

In a pairwise comparison of KRAS G12/G13 mutant versus KRAS G12/G13 wild-type, a statistically significant improvement in disease-specific survival was observed in the KRAS G12/G13 wild-type cohort (p-value = 0.02, log rank).

Supplemental Figure 3. Frequency and concordance of KRAS, NRAS and BRAF mutations in the colorectal cancer cell lines. A. Genomic DNA was extracted from 70 colorectal cancer cell lines and was screened for mutations in *KRAS*, *NRAS*, *HRAS*, *BRAF*, *MEK1* and *PIK3CA* using the MALDI-TOF based genotyping platform. Exon 2 KRAS mutations (G12 and G13) are shown in dark blue, exon 3 KRAS mutations (Q61) in light blue, exon 4 KRAS mutations (K117 and A146) in red, and BRAF mutations in green. **B.** Exon 4 KRAS mutations (all either K117N or A146T/V) were non-overlapping in distribution with mutations in KRAS exons 2 and 3 and BRAF. No colorectal cancer cell lines with NRAS mutation were identified. **C.** Graphical representation of the hotspot mutations in KRAS as mapped to the *KRAS4b* gene. Introns are represented as thin lines whereas exons (not to scale) are shown as boxes with the blue shading representing the coding regions and the green shading representing the untranslated regions (UTR). Red dots below the exons represent sites for hotspot mutations found in the human tumors and cell lines.