# Supplementary Information

**Supplementary Information Table 1 | *Kras* mutation status of carcinogen-induced mouse lung adenomas**

| Tumour | Treatment | Strain | Grade | Genotype | *Kras* status (WES)* | *Kras* status (Sanger)† |
|---|---|---|---|---|---|---|
| 32T1 | Urethane | FVB | Adenoma | Kras+/- | Q61L | Q61L |
| 32T2 | Urethane | FVB | Adenoma | Kras+/- | Q61R | Q61R |
| 32T3 | Urethane | FVB | Adenoma | Kras+/- | Q61L | Q61L |
| 33T1 | Urethane | FVB | Adenoma | Kras+/- | Q61L | Q61L |
| 33T2 | Urethane | FVB | Adenoma | Kras+/- | Q61L | Q61L |
| 33T3 | Urethane | FVB | Adenoma | Kras+/- | Q61L | Q61L |
| 33T4 | Urethane | FVB | Adenoma | Kras+/- | Q61L | Q61L |
| 33T5 | Urethane | FVB | Adenoma | Kras+/- | Q61L | Q61L |
| 47T2 | Urethane | FVB | Adenoma | Kras+/- | Q61L | Q61L |
| 48T2 | Urethane | FVB | Adenoma | Kras+/- | Q61L | Q61L |
| 48T3 | Urethane | FVB | Adenoma | Kras+/- | Q61L | Q61L |
| 48T5 | Urethane | FVB | Adenoma | Kras+/- | Q61L | Q61L |
| 1793T1 | Urethane | FVB | Adenoma | Kras+/- | Q61L | Q61L |
| 1793T2 | Urethane | FVB | Adenoma | Kras+/- | Q61L | Q61L |
| 1793T3 | Urethane | FVB | Adenoma | Kras+/- | Q61L | Q61L |
| 1800T1 | Urethane | FVB | Adenoma | Kras+/- | Q61L | Q61L |
| 1800T2 | Urethane | FVB | Adenoma | Kras+/- | Q61L | Q61L |
| 1800T3 | Urethane | FVB | Adenoma | Kras+/- | Q61L | Q61L |
| 309T1 | Urethane | FVB | Adenoma | Kras+/- | Q61L | Q61L |
| 309T3 | Urethane | FVB | Adenoma | Kras+/- | Q61L | Q61L |
| 66T1 | Urethane | FVB | Adenoma | Kras+/- | Q61L | Q61L |
| 66T2 | Urethane | FVB | Adenoma | Kras+/- | Q61L | Q61L |
| 66T3 | Urethane | FVB | Adenoma | Kras+/- | Q61H | Q61H |
| 75T1 | Urethane | FVB | Adenoma | Kras+/- | Q61L | Q61L |
| 75T2 | Urethane | FVB | Adenoma | Kras+/- | Q61L | Q61L |
| 75T3 | Urethane | FVB | Adenoma | Kras+/- | Q61L | Q61L |
| 45T3 | Urethane | FVB | Adenoma | WT | Q61R | NA |
| 45T4 | Urethane | FVB | Adenoma | WT | Q61R | No mut |
| 45T5 | Urethane | FVB | Adenoma | WT | Q61R | Q61R |
| 45T6 | Urethane | FVB | Adenoma | WT | Q61R | Q61R |
| 44T2 | Urethane | FVB | Adenoma | WT | Q61R | Q61R |
| 44T3 | Urethane | FVB | Adenoma | WT | Q61R | Q61R |
| 35T1 | Urethane | FVB | Adenoma | WT | Q61R | Q61R |
| 35T2 | Urethane | FVB | Adenoma | WT | Q61R | Q61R |
| 35T3 | Urethane | FVB | Adenoma | WT | Q61R | Q61R |
| 63T1 | Urethane | FVB | Adenoma | WT | Q61R | NA |
| 1T1 | Urethane | FVB | Adenoma | WT | Q61L | Q61L |
| 1T2 | Urethane | FVB | Adenoma | WT | Q61R | NA |
| 1T3 | Urethane | FVB | Adenoma | WT | Q61R | No mut |
| 2T1 | Urethane | FVB | Adenoma | WT | Q61R | Q61R |
| 2T2 | Urethane | FVB | Adenoma | WT | Q61R | Q61R |
| 2T3 | Urethane | FVB | Adenoma | WT | Q61R | Q61R |
| 1790T1 | Urethane | FVB | Adenoma | WT | Q61R | Q61R |
| 1790T3 | Urethane | FVB | Adenoma | WT | Q61R | Q61R |
| 1039T2 | MNU | FVB | Adenoma | Kras+/- | G12D | NA |
| 1039T3 | MNU | FVB | Adenoma | Kras+/- | G12D | G12D |
| 1039T4 | MNU | FVB | Adenoma | Kras+/- | G12D | G12D |
| 1039T5 | MNU | FVB | Adenoma | Kras+/- | G12D | NA |
| 1024T1 | MNU | FVB | Adenoma | Kras+/- | G12D | G12D |
| 1024T2 | MNU | FVB | Adenoma | Kras+/- | G12D | G12D |
| 1024T3 | MNU | FVB | Adenoma | Kras+/- | G12D | G12D |
| 1024T4 | MNU | FVB | Adenoma | Kras+/- | G12D | G12D |
| 1024T5 | MNU | FVB | Adenoma | Kras+/- | G12D | NA |
| 1024T6 | MNU | FVB | Adenoma | Kras+/- | G12D | NA |
| 1024T7 | MNU | FVB | Adenoma | Kras+/- | G12D | G12D |
| 1024T8 | MNU | FVB | Adenoma | Kras+/- | G12D | G12D |
| 1024T9 | MNU | FVB | Adenoma | Kras+/- | G12D | G12D |
| 1024T10 | MNU | FVB | Adenoma | Kras+/- | G12D | G12D |
| 1024T11 | MNU | FVB | Adenoma | Kras+/- | G12D | G12D |
| 1045T1 | MNU | FVB | Adenoma | Kras+/- | G12D | No mut |
| 1045T2 | MNU | FVB | Adenoma | Kras+/- | G12D | G12D |
| 1045T3 | MNU | FVB | Adenoma | Kras+/- | G12D | NA |
| 1045T4 | MNU | FVB | Adenoma | Kras+/- | G12D | G12D |
| 1045T5 | MNU | FVB | Adenoma | Kras+/- | G12D | G12D |
| 1045T7 | MNU | FVB | Adenoma | Kras+/- | G12D | G12D |
| 1012T3 | MNU | FVB | Adenoma | WT | No mut | NA |
| 1026T1 | MNU | FVB | Adenoma | WT | G12D | G12D |
| 1026T2 | MNU | FVB | Adenoma | WT | G12D | NA |
| 1011T1 | MNU | FVB | Adenoma | WT | G12D | NA |
| 1029T1 | MNU | FVB | Adenoma | WT | G12D | NA |

† NA = data not available due to assay failure

* *Kras* status (WES): *Kras* mutation determined by whole-exome sequencing

**Supplementary Information Table 2 | Mouse lung adenoma cancer driver gene SNVs compared against COSMIC database of human cancers**

This table is provided as a separate excel attachment (SI_Table2.xlxs) in Supplementary Information.

**Supplementary Information Table 3 | *MTUS1* expression and overall survival in lung cancer patients**

| Study | Platform | Class | Stages | Patients (n) | Covariates | χ2 | p-value |
|---|---|---|---|---|---|---|---|
| TCGA_LUAD exp_HiSeqV2 | Illumina HiSeq 2000 RNA Seq | ADCA | I-IV | 354 | Sex, Age, Pack Years, Stage | 10.9 | 0.00097 |
| Shedden K, et al. (2008) | Affymetrix U133A Expression | ADCA | I-III | 440 | Sex, Age, Stage | 4.4 | 0.036 |
| TCGA_LUSC exp_HiSeqV2 | Illumina HiSeq 2000 RNA Seq | SCC | I-IV | 308 | Sex, Age, Pack Years, Stage | 0.1 | 0.76 |
| Raponi M, et al. (2006) | Affymetrix U133A Expression | SCC | I-III | 130 | Sex, Age, Pack Years, Stage | 1.83 | 0.18 |

**Supplementary Information Table 4 | *Kras* mutation status of urethane-induced mouse lung adenocarcinomas**

| Tumour | Treatment | Strain | Grade* | Genotype | *Kras* status (WES)† | *Kras* status (Sanger) |
|---|---|---|---|---|---|---|
| 006B2 AC1 | Urethane | A/J | Adenocarcinoma | WT | NA | Q61R |
| 006B3 AC | Urethane | A/J | Adenocarcinoma | WT | G12V | G12V |
| 006C8 AC1 | Urethane | A/J | Adenocarcinoma | WT | NA | Q61L |
| 006C8 AC2 | Urethane | A/J | Adenocarcinoma | WT | Q61R | Q61R |
| 004E3 AC1 | Urethane | A/J | Adenocarcinoma | WT | Q61R | Q61R |
| 004E3 AC2 | Urethane | A/J | Adenocarcinoma | WT | Q61L | Q61L |
| 004E4 AC1 | Urethane | A/J | Adenocarcinoma | WT | Q61R | Q61R |
| 004E4 AC2 | Urethane | A/J | Adenocarcinoma | WT | Q61R | Q61R |
| 004E4 AC3 | Urethane | A/J | Adenocarcinoma | WT | NA | Q61R |
| 011A5 AC | Urethane | A/J | Adenocarcinoma | WT | NA | Q61R |
| 010E7 AC1 | Urethane | A/J | NA | WT | Q61R | Q61R |
| 010E7 AC4 | Urethane | A/J | NA | WT | Q61R | Q61R |
| 010E7 AC6 | Urethane | A/J | NA | WT | NA | Q61R |
| 51 AC1 | Urethane | FVB | Adenocarcinoma | Kras+/- | Q61L | Q61L |
| 66 AC1 | Urethane | FVB | Adenocarcinoma | WT | Q61R | Q61R |
| 2913 AC1 | Urethane | FVB | Adenocarcinoma | Kras+/- | Q61L | Q61L |
| 2913 AC2 | Urethane | FVB | Adenocarcinoma | Kras+/- | Q61L | Q61L |
| 2915 AC1 | Urethane | FVB | Adenocarcinoma | Kras+/- | Q61L | Q61L |
| 2915 AC2 | Urethane | FVB | Adenocarcinoma | Kras+/- | Q61L | Q61L |
| 2916 AC1 | Urethane | FVB | Adenocarcinoma | Kras+/- | Q61H | Q61H |
| 2916 AC2 | Urethane | FVB | Adenocarcinoma | Kras+/- | Q61L | Q61L |
| 2916 AC3 | Urethane | FVB | Adenocarcinoma | Kras+/- | Q61L | Q61L |

* NA = data not available due to insufficient tissue for histology

† *Kras* status (WES): *Kras* mutation determined by whole-exome sequencing. NA = data not available due to poor coverage at *Kras* in some samples

**Supplementary Information Table 5 | Mouse adenocarcinoma SNVs in established cancer driver genes and *Fat4***

| Chr | Position | Gene | Exon | Substitution | Consequence | Observed | Tumours | Validated | Validation Method |
|---|---|---|---|---|---|---|---|---|---|
| 2 | 122151067 | B2m | 2 | CAC>G | T88A | 1 | 004E3 AC1 | Yes | Sanger |
| 3 | 84904008 | Fbxw7 | 3 | CAC>T | T147S | 1 | 006B3 AC | Yes | Sanger |
| 6 | 47545835 | Ezh2 | 10 | AGG>A | G356R | 1 | 2916 AC3 | Yes | Sanger |
| 6 | 113624325 | Vhl | 1 | CAG>C | Q62P | 1 | 006B3 AC | Yes | Sanger |
| 9 | 110594639 | Setd2 | 15 | AGT>A | S2254N | 1 | 2915 AC1 | Yes | Sanger |
| 10 | 81531160 | Gna11 | 6 | AAC>G | N274S | 1 | 010E7 AC1 | Yes | Sanger |
| 11 | 4782278 | Nf2 | 15 | AAA>G | K550R | 1 | 004E3 AC2 | Yes | Sanger |
| 12 | 3896970 | Dnmt3a | 5 | GAA>G | N180D | 1 | 66 AC1 | Yes | Sanger |
| 13 | 111750558 | Map3k1 | 19 | CAG>T | Q1406L | 1 | 2913 AC1 | Yes | Sanger |
| 15 | 98857427 | Mll2 | 19 | CGC>A | A1524T | 1 | 006C8 AC1 | Yes | Sanger |
| 19 | 6335662 | Men1 | 2 | CGC>A | A74T | 1 | 2913 AC1 | Yes | Sanger |
| 3 | 38887878 | Fat4 | 1 | CGT>A | V307M | 1 | 66 AC1 | Yes | Sanger |
| 3 | 38951175 | Fat4 | 4 | AAC>G | T1908A | 1 | 2916 AC2 | Yes | Sanger |
| 3 | 38981249 | Fat4 | 9 | AAT>C | L3017V | 1 | 006B3 AC | Yes | Sanger |

**Supplementary Information Table 6 | Comparison of mutations involving driver genes in all carcinogen-induced mouse lung tumours and human lung adenocarcinoma**

| Gene | LUAD obs | LUAD % | *KRAS* LUAD obs | *KRAS* LUAD % | Mouse obs | Mouse % |
|---|---|---|---|---|---|---|
| AKT1 | 2 | 0.9 | 1 | 1.5 | 1 | 1.1 |
| ALK | 19 | 8.3 | 5 | 7.4 | 1 | 1.1 |
| APC | 9 | 3.9 | 2 | 2.9 | 1 | 1.1 |
| ARID1A | 17 | 7.4 | 5 | 7.4 | 2 | 2.2 |
| ARID1B | 10 | 4.3 | 5 | 7.4 | 3 | 3.3 |
| ASXL1 | 4 | 1.7 | 0 | 0 | 2 | 2.2 |
| ATM | 25 | 10.9 | 13 | 19.1 | 3 | 3.3 |
| ATRX | 15 | 6.5 | 3 | 4.4 | 1 | 1.1 |
| AXIN1 | 2 | 0.9 | 1 | 1.5 | 1 | 1.1 |
| B2M | 1 | 0.4 | 0 | 0 | 1 | 1.1 |
| BRCA1 | 7 | 3 | 2 | 2.9 | 1 | 1.1 |
| BRCA2 | 11 | 4.8 | 4 | 5.9 | 1 | 1.1 |
| CARD11 | 14 | 6.1 | 4 | 5.9 | 1 | 1.1 |
| CBL | 4 | 1.7 | 1 | 1.5 | 1 | 1.1 |
| CIC | 3 | 1.3 | 1 | 1.5 | 2 | 2.2 |
| CREBBP | 10 | 4.3 | 4 | 5.9 | 3 | 3.3 |
| DAXX | 2 | 0.9 | 1 | 1.5 | 1 | 1.1 |
| DNMT3A | 6 | 2.6 | 3 | 4.4 | 2 | 2.2 |
| EP300 | 2 | 0.9 | 1 | 1.5 | 1 | 1.1 |
| EZH2 | 4 | 1.7 | 1 | 1.5 | 1 | 1.1 |
| FAM123B | 14 | 6.1 | 6 | 8.8 | 1 | 1.1 |
| FBXW7 | 4 | 1.7 | 3 | 4.4 | 1 | 1.1 |
| FGFR2 | 5 | 2.2 | 1 | 1.5 | 1 | 1.1 |
| FGFR3 | 1 | 0.4 | 1 | 1.5 | 2 | 2.2 |
| FGFR4 | 3 | 1.3 | 0 | 0 | 1 | 1.1 |
| FLT3 | 11 | 4.8 | 1 | 1.5 | 1 | 1.1 |
| GATA2 | 1 | 0.4 | 1 | 1.5 | 1 | 1.1 |
| GATA3 | 6 | 2.6 | 1 | 1.5 | 1 | 1.1 |
| GNA11 | 0 | 0 | 0 | 0 | 1 | 1.1 |
| HNF1A | 0 | 0 | 0 | 0 | 1 | 1.1 |
| IDH1 | 3 | 1.3 | 2 | 2.9 | 1 | 1.1 |
| JAK2 | 9 | 3.9 | 6 | 8.8 | 1 | 1.1 |
| KDM5C | 5 | 2.2 | 2 | 2.9 | 2 | 2.2 |
| KIT | 5 | 2.2 | 3 | 4.4 | 1 | 1.1 |
| KLF4 | 3 | 1.3 | 0 | 0 | 1 | 1.1 |
| MAP3K1 | 4 | 1.7 | 2 | 2.9 | 2 | 2.2 |
| MAP3K9 | 6 | 2.6 | 2 | 2.9 | 3 | 3.3 |
| MED12 | 16 | 7 | 6 | 8.8 | 1 | 1.1 |
| MEN1 | 1 | 0.4 | 1 | 1.5 | 2 | 2.2 |
| MET | 17 | 7.4 | 3 | 4.4 | 1 | 1.1 |
| MLL2 | 19 | 8.3 | 8 | 11.8 | 3 | 3.3 |
| MTUS1 | 4 | 1.7 | 0 | 0 | 4 | 4.3 |
| NCOR1 | 11 | 4.8 | 3 | 4.4 | 1 | 1.1 |
| NF1 | 30 | 13 | 1 | 1.5 | 1 | 1.1 |
| NF2 | 1 | 0.4 | 0 | 0 | 1 | 1.1 |
| NOTCH2 | 7 | 3 | 2 | 2.9 | 1 | 1.1 |
| PAX5 | 1 | 0.4 | 0 | 0 | 1 | 1.1 |
| PDGFRA | 15 | 6.5 | 6 | 8.8 | 2 | 2.2 |
| PPP2R1A | 4 | 1.7 | 0 | 0 | 2 | 2.2 |
| PTCH1 | 11 | 4.8 | 4 | 5.9 | 1 | 1.1 |
| RB1 | 10 | 4.3 | 2 | 2.9 | 6 | 6.5 |
| RET | 9 | 3.9 | 3 | 4.4 | 1 | 1.1 |
| RNF43 | 3 | 1.3 | 0 | 0 | 1 | 1.1 |
| SETD2 | 20 | 8.7 | 4 | 5.9 | 1 | 1.1 |
| SF3B1 | 5 | 2.2 | 0 | 0 | 1 | 1.1 |
| TNFAIP3 | 2 | 0.9 | 1 | 1.5 | 1 | 1.1 |
| VHL | 0 | 0 | 0 | 0 | 1 | 1.1 |
| | **Tumours = 230** | | **Tumours = 68** | | **Tumours = 92** | |

Number and % of consequential mutations in TCGA LUAD WES

Number and % of consequential mutations in TCGA LUAD *KRAS*-mutant tumours

Number and % of consequential mutations in carcinogen-induced mouse tumours in this study

**Supplementary Information Table 7 | Mouse and human SNVs of cancer driver genes occurring in close proximity**

| Gene | Mouse mut | Mouse obs | Human hom aa pos | LUAD mut | LUAD obs | Mouse SNV Consequence |
|---|---|---|---|---|---|---|
| AKT1 | E40K | 1 | 40 | E17K | 1 | Constitutive activation |
| ALK | G133R | 1 | 129 | V132L | 1 | |
| ARID1A | S520F | 1 | 519 | Q479* | 1 | |
| ARID1B | P564L | 1 | 554 | G530C | 1 | |
| ASXL1 | P430S | 1 | 430 | G434V | 1 | |
| ATM | S367F | 1 | 367 | Q355_splice | 1 | Autophosphorylation site: loss of function |
| ATM | E648K | 1 | 649 | Q628* | 1 | |
| B2M | T88A | 1 | 88 | E70* | 1 | |
| CBL | S401F | 1 | 403 | C401F, D429_del | 2 | RING finger domain hotspot |
| CIC | I1432F | 1 | 1436 | E1405* | 1 | |
| CREBBP | P1353S | 1 | 1352 | K1367N | 1 | |
| DAXX | D596N | 1 | 595 | L566F | 1 | |
| EZH2 | G356R | 1 | 361 | E341K | 1 | |
| FBXW7 | T147S | 1 | 144 | E113* | 1 | |
| GATA2 | Y376C | 1 | 376 | S340_splice | 1 | |
| MEN1 | G169R | 1 | 163 | A165D | 1 | |
| MET | K1196Q | 1 | 1216 | Y1235H | 1 | |
| NF2 | K550R | 1 | 550 | E527Q | 1 | |
| NOTCH2 | P426S | 1 | 426 | S462* | 1 | |
| PDGFRA | N711I | 1 | 711 | E726Q | 1 | |
| PDGFRA | D877N | 1 | 877 | A916T | 1 | |
| PPP2R1A | P523L | 1 | 523 | A541G | 1 | |
| RB1 | S744F, S766F | 2 | 751, 773 | Y756C | 1 | |

Table of carcinogen-induced mouse SNVs and human TCGA LUAD SNVs in cancer driver genes occurring within + or - 40 aa of homologous position

**Supplementary Information Table 8 | Adenocarcinoma CGN>A (NCG>T) signature validation**

| Chr | Position | Gene | Exon* | Substitution | Consequence | Obs | Tumours | Validated† | Validation Method‡ |
|---|---|---|---|---|---|---|---|---|---|
| 15 | 98857427 | Mll2 | 19 | CGC>A | A1524T | 1 | 006C8 AC1 | Yes | Sanger |
| 19 | 6335662 | Men1 | 2 | CGC>A | A74T | 1 | 2913 AC1 | Yes | Sanger |
| 3 | 38887878 | Fat4 | 1 | CGT>A | V307M | 1 | 66 AC1 | Yes | Sanger |
| 19 | 6255487 | Atg2a | 27 | CGC>A | A1266T | 1 | 006C8 AC1 | Yes | Sanger |
| 1 | 53851676 | Hecw2 | 21 | CGG>A | R1227X | 1 | 006C8 AC1 | Yes | Sanger |
| 1 | 135402359 | Ipo9 | 13 | CGA>A | E448K | 1 | 2913 AC1 | Yes | Sanger |
| 4 | 3748753 | Lyn | 8 | CGG>A | R244W | 1 | 004E3 AC2 | Yes | Sanger |
| 8 | 41000849 | Mtus1 | 9 | CGG>A | A982A | 1 | 2916 AC2 | Yes | Sanger |
| 15 | 76189287 | Plec | 13 | CGG>A | R306Q | 1 | 004E3 AC2 | Yes | Sanger |
| 17 | 86493271 | Prkce | 9 | CGG>A | R372X | 1 | 2915 AC1 | Yes | Sanger |
| 14 | 51130312 | Rnase6 | 2 | CGC>A | R54C | 1 | 66 AC1 | Yes | Sanger |
| 16 | 35663231 | Sema5b | 22 | CGG>A | R1080W | 1 | 004E4 AC3 | Yes | Sanger |
| 17 | 80406840 | Sos1 | 20 | CGG>A | R1067Q | 1 | 2913 AC2 | Yes | Sanger |
| 2 | 21746709 | Gpr158 | 6 | CGA>A | R480C | 1 | 006C8 AC2 | Yes | Sanger |
| 10 | 38966081 | Lama4 | 1 | CGG>A | G43S | 1 | 2916 AC1 | Yes | Both |
| 12 | 54068154 | Npas3 | 12 | CGG>A | R620X | 1 | 004E4 AC1 | Yes | Visual inspection |
| 8 | 83424085 | Clgn | 12 | CGG>A | R460X | 1 | 006B2 AC1 | Inconclusive | Both |
| 10 | 94514984 | Tmcc3 | NA | CGC>A | NA | 1 | 2913 AC2 | Inconclusive | Both |
| 11 | 3132472 | Sfi1 | NA | CGT>A | NA | 1 | 2913 AC1 | Inconclusive | Both |
| 12 | 57542962 | Foxa1 | 2 | CGC>A | G157G | 1 | 51 AC1 | Inconclusive | Both |
| 15 | 80692611 | A430088P11Rik | NA | CGC>A | NA | 1 | 2915 AC1 | Inconclusive | Both |
| 19 | 53618124 | Smc3 | 5 | CGA>A | S85L | 1 | 2913 AC1 | Inconclusive | Both |
| X | 73778326 | Srpk3 | 14 | CGG>A | R498Q | 1 | 2916 AC3 | Inconclusive | Both |

* NA = non-exonic location.

† Variants were called inconclusive if visual inspection supported somatic origin, but variant read depth was <10.

‡ Validation Method: Both = Sanger sequencing and visual inspection.