Supplementary Table 1. Advances in carcinogenesis: description and references used for Figure 1

Supplementary Table 2. Summary and references of the data used for each analysis.

Supplementary Figure 1A and B. Detailed mutation spectrum observed in lung cancer

Supplementary Figure 2A and B. Detailed mutation spectrum observed in colorectal and glioblastoma cancer. G:C->T:A transitions at CpG sites are shown in red when data are available.

Supplementary Figure 3A and B. Detailed mutation spectrum observed in breast cancer. G:C->T:A transitions at CpG sites are shown in red when data are available.

Supplementary Figure 4A. Detailed mutation spectrum observed in hepatocellular carcinoma (HCC). AflB Neg: patients not exposed to Aflatoxin B1; AflB Pos: patients exposed to Aflatoxin B1; Europe: Patients from European countries (exposed and not exposed); China: Patients from China (exposed and not exposed); RIKEN and NCC: mutational data for liver cancer from the International Cancer Genome Consortium (http://dcc.icgc.org/) obtained from two Japanese institutions, RIKEN and the National Cancer Center.

Supplementary Figure 5. Detailed mutation spectrum observed in prostate cancer. Data were obtained from the UMD TP53 database (TP53) and for patient 1 to 7 (Pat 1-7) from Berger et al (2011) The genomic complexity of primary human prostate cancer. Nature 470: 214-220.

Supplementary Figure 6. Detailed mutation spectrum observed in ovarian cancer. Data were obtained from the UMD TP53 database (TP53) and from Bell et al. (2011) Integrated genomic analyses of ovarian carcinoma. *Nature* **474**, 609-615.

Supplementary Figure 7: Full size of publication's figure 2.
Supplementary Figure 8: Full size of publication's figure 3.
Supplementary Figure 9: Full size of publication's figure 4.
Supplementary Figure 10: Full size of publication's figure 5.
Supplementary Figure 11: Full size of publication's figure 6.
Supplementary Figure 12: Full size of publication's figure 7.
Supplementary Figure 13: Full size of publication's figure 8.

Event	Date	References
First description of an association between occupational exposure and lung cancer (Paracelsus)	1567	(Theophrastus Bombastus von Hohenheim (Paracelsus), 1567)
Publication of the first book on occupational diseases and industrial hygiene (Ramazzini)	1713	(Ramazzini, 1700)
Association between nasal cancer and snuff consumption (Hill)	1761	(Hill, 1761)
Association between scrotal cancer and chimney sweeps (P. Pott)	1775	(Pott, 1775)
Association between arsenical vapor and skin cancer (Ayrton)	1822	(Paris, 1822)
Association between cancer incidence and age (Rigoni-Stern)	1842	(Rigoni-Stern, 1842)
Association between UV exposure and skin cancer (Thiersch)	1875	(Thiersch, 1875)
High frequency of bladder cancer in aniline workers in dyestuff factories (Rehn)	1895	(Rehn, 1895)
Chromosomal mutations in cancer	1914	(Boveri, 1914)
First experimental cancer produced by coal tar on rabbits (Yamagiwa)	1915	(Yamagiwa and Ichikawa, 1918)
Induction of skin cancer in mice by arsenic	1922	(Leitch and Kennaway, 1922)
Tumors exhibit anaerobic respiration	1924	(Warburg et al., 1924)
First proof that single polycyclic aromatic hydrocarbons (PAHs) are capable of inducing malignant skin tumors in mice	1930	(Kennaway and Hieger, 1930)
Synthesis of benzo[a]pyrene	1930	(Kennaway, 1930)
Isolation of benzo[a]pyrene from coal tar and proof of its carcinogenicity in mouse skin	1933	(Cook et al., 1933)
benzo[a]pyrene produces cancer in mouse skin	1933	(Cook et al., 1933)
The tricyclic aromatic hydrocarbon anthracene is hydroxylated *in vivo*	1935	(Boyland and Levi, 1935)
Anthracene metabolized *in vivo*	1935	(Boyland and Levi, 1935)
2-naphthylamine induces papillomas	1938	(Hueper et al., 1938)
Production of bladder papillomas and carcinomas in dogs by 2-naphthylamine	1938	(Hueper et al., 1938)
Two-stage carcinogenesis: initiation promotion	1941	(Berenblum, 1941)
Initiating and promoting effects of chemical carcinogens are distinguished	1944	(Friedewald and Rous, 1944)
Covalent binding of N,N-dimethyl-4-aminoazobenzene to proteins in rat liver	1947	(Miller and Miller, 1947)
Microsome-catalysed biotransformation of N,N-dimethyl-4-aminoazobenzene in a cell-free system 42	1948	(Mueller and Miller, 1948)
Association between smoking and lung cancer (Wynder)	1950	(Doll and Hill, 1950; Wynder and Graham,

		1950)
Discovery of DNA structure (Watson and Crick)	1953	(Watson and Crick, 1953)
N-nitrosamine methylate DNA	1956	(Magee and JM, 1956)
Enzyme induction by polycyclic aromatic hydrocarbons	1956	(Conney et al., 1956)
First evidence that epoxides are intermediates in biotransformation of PAHs	1960	(Boyland and Sims, 1960)
Ultraviolet induces pyrimidine dimer in DNA	1961	(Setlow and Setlow, 1961)
Cyt P450 is localized in microsomes	1962	(Omura and Sato, 1962)
In vitro transformation of cells by chemical carcinogens	1963	(Berwald and Sachs, 1963)
Induction of transcription by methycholanthrene	1963	(Loeb and Gelboin, 1963)
Correlation between DNA-binding level and carcinogenicity of six PAHs	1964	(Brookes and Lawley, 1964)
Aflatoxin B1 is carcinogenic in rats	1964	(Sporn et al., 1966)
Aflatoxin binds to DNA	1966	(Sporn et al., 1966)
Aflatoxin induces liver cancer	1968	(Alpert et al., 1968)
DNA repair and cancer	1968	(Cleaver, 1968)
Retinoblastoma requires two mutations	1971	(Knudson, 1971)
Endogenous DNA damage	1972	(Lindahl and Nyberg, 1972)
Ames assay for chemical mutagens	1973	(Ames et al., 1973)
Benzo[a]pyrene binds to DNA through its 7,8-diol-9,10-epoxide	1974	(Sims et al., 1974)
Mutator phenotype	1974	(Loeb et al., 1974)
Malonaldehyde is carcinogenic	1974	(Shamberger et al., 1974)
Chemical carcinogens are activated to form DNA adducts in human tissue	1974	(Harris et al., 1974)
Stereoselective enzymatic conversion of BP leads to the 7R,8S-diol-9S,10R-epoxide, (+)-anti-BP diol-epoxide (BPDE)	1976	(Yang et al., 1976)
Clonal selection of mutant cancer cells	1976	(Nowell, 1976)
Interindividual variation in carcinogen activation	1976	(Harris et al., 1976)
Stereospecificity of benzo[a]pyrene diolepoxide	1976	(Yang et al., 1976)
Mutagens and carcinogens in cooked food	1977	(Nagao et al., 1977)
Binding of benzo[a]pyrene to DNA *in vivo* predominantly occurs via (+)-anti-BPDE	1978	(Koreeda et al., 1978)
Tobacco-specific N-nitrosamine	1978	(Hecht et al., 1978)
TP53 discovery	1979	(DeLeo et al., 1979; Kress et al., 1979; Lane and Crawford, 1979; Linzer and Levine, 1979)
Landmark study that shows epidemiological evidence for a predominent role of environmental factors in human cancer	1981	(Doll and Peto, 1981)
Benzo[a]pyrene formed adducts in human DNA	1982	(Perera et al., 1982)

Induction of activating Ha-ras mutations in mouse skin following exposure to PAHs	1983	(Shih and Weinberg, 1982; Balmain and Pragnell, 1983)
Mutations in Ha-ras following exposure to carcinogens	1983	(Shih and Weinberg, 1982; Balmain and Pragnell, 1983)
8-oxo-dG as a marker for oxidative damage	1986	(Grollman and Moriya, 1993)
The exo-8,9-epoxide is the DNA-binding metabolite of aflatoxin B	1988	(Baertschi et al., 1988)
Model for cancer progression	1990	(Fearon and Vogelstein, 1990)
Mutation spectrum reflects carcinogen exposure	1991	(Hollstein et al., 1991; Caron de Fromentel and Soussi, 1992)
TP53 mutation in skin cancer correlates with UV exposure	1991	(Brash et al., 1991)
Aflatoxin induces a specific mutation at codon 249 of the TP53 gene	1991	(Bressac et al., 1991; Hsu et al., 1991)
Selective targeting of p53 gene mutational hotspots in human cancers by Aflatoxin B1 and BaP.	1991	(Puisieux et al., 1991)
UV photoproducts at mutational hotspot of the TP53 gene specific for skin cancer.	1993	(Tornaletti et al., 1993)
DNA-binding signature of BPDE in TP53 of lung epithelial cells corresponds to human lung cancer mutational hot spots	1996	(Denissenko et al., 1996)
MSI tumors in HNPCC	1996	
Translesion DNA synthesis	1999	
Mutation in single cancer cells	1999	
Cancer genome contains multiple driver and passenger mutations	2000	(Haber and Settleman, 2007)
UV-induced DNA damage and mutations in Hupki mice recapitulate p53 hotspot alterations in skin cancer	2001	(Luo et al., 2001)
p53 mutations in benzo(a)pyrene-exposed hupki mice fibroblasts correlate with p53 mutations in human lung tumors.	2005	(Liu et al., 2005)
Cancer genome mutations reflect mutagen exposure	2007	(Davies et al., 2005; Greenman et al., 2007)
First release of a cancer genome sequence (ALL)	2008	(Ley et al., 2008)
Several cancer genome sequences are released	2009	(Shah et al., 2009) (Ding et al., 2010) (Pleasance et al., 2010a; Pleasance et al., 2010b)
Multiple tumor genome from the same type of cancer are fully sequenced and compared	2011	(Berger et al., 2011; Chapman et al., 2011)

Alpert, M.E., M.S. Hutt and C.S. Davidson. 1968. Hepatoma in Uganda. A study in geographic pathology. *Lancet* 1: 1265-1267.

Ames, B.N., W.E. Durston, E. Yamasaki and F.D. Lee. 1973. Carcinogens are mutagens: a simple test system combining liver homogenates for activation and bacteria for detection. *Proc Natl Acad Sci U S A* 70: 2281-2285.

Baertschi, W.S., K.D. Raney, M.P. Stone and T.M. Harris. 1988. Preparation of the 8,9-epoxide of the mycotoxin aflatoxin B1: The ultimate carcinogenic species. *Journal of the American Chemical Society* 110: 7929-7931.

Balmain, A. and I.B. Pragnell. 1983. Mouse skin carcinomas induced in vivo by chemical carcinogens have a transforming Harvey-ras oncogene. *Nature* 303: 72-74.

Berenblum, I. 1941. The Cocarcinogenic Action of Croton Resin. *Cancer Res* 1: 44-48.

Berger, M.F. et al. 2011. The genomic complexity of primary human prostate cancer. *Nature* 470: 214-220.

Berwald, Y. and L. Sachs. 1963. In vitro cell transformation with chemical carcinogens. *Nature* 200: 1182-1184.

Boveri, T. 1914. Zur Frage der Entstehung maligner Tumoren. *Jena: Gustave Fischer Verlag*
Boyland, E. and A.A. Levi. 1935. Metabolism of polycyclic compounds: Production of dihydroxydihydroanthracene from anthracene. *Biochem J* 29: 2679-2683.

Boyland, E. and P. Sims. 1960. Metabolism of polycyclic compounds. 16. The metabolism of 1:2-dihydronaphthalene and 1:2-epoxy-1:2:3:4-tetrahydronaphthalene. *Biochem J* 77: 175-181.

Brash, D.E., J.A. Rudolph, J.A. Simon, A. Lin, G.J. McKenna, H.P. Baden, A.J. Halperin and J. Ponten. 1991. A role for sunlight in skin cancer: UV-induced p53 mutations in squamous cell carcinoma. *Proc Natl Acad Sci U S A* 88: 10124-10128.

Bressac, B., M. Kew, J. Wands and M. Ozturk. 1991. Selective G to T mutations of p53 gene in hepatocellular carcinoma from southern Africa. *Nature* 350: 429-431.

Brookes, P. and P.D. Lawley. 1964. Evidence for the binding of polynuclear aromatic hydrocarbons to the nucleic acids of mouse skin: relation between carcinogenic power of hydrocarbons and their binding to deoxyribonucleic acid. *Nature* 202: 781-784.

Caron de Fromentel, C. and T. Soussi. 1992. TP53 tumor suppressor gene: a model for investigating human mutagenesis. *Genes Chromosomes Cancer* 4: 1-15.

Chapman, M.A. et al. 2011. Initial genome sequencing and analysis of multiple myeloma. *Nature* 471: 467-472.

Cleaver, J.E. 1968. Defective repair replication of DNA in xeroderma pigmentosum. *Nature* 218: 652-656.

Conney, A.H., E.C. Miller and J.A. Miller. 1956. The metabolism of methylated aminoazo dyes. V. Evidence for induction of enzyme synthesis in the rat by 3-methylcholanthrene. *Cancer Res* 16: 450-459.

Cook, J.W., C.L. Hewett and I. Hieger. 1933. The isolation of a cancer-producing hydrocarbon from coal tar. Parts I, II, and III. *Chem Soc* 24: 395-405.

Davies, H. et al. 2005. Somatic mutations of the protein kinase gene family in human lung cancer. *Cancer Res* 65: 7591-7595.

DeLeo, A.B., G. Jay, E. Appella, G.C. Dubois, L.W. Law and L.J. Old. 1979. Detection of a transformation-related antigen in chemically induced sarcomas and other transformed cells of the mouse. *Proc Natl Acad Sci U S A* 76: 2420-2424.

Denissenko, M.F., A. Pao, M. Tang and G.P. Pfeifer. 1996. Preferential formation of benzo[a]pyrene adducts at lung cancer mutational hotspots in P53. *Science* 274: 430-432.

Ding, L. et al. 2010. Genome remodelling in a basal-like breast cancer metastasis and xenograft. *Nature* 464: 999-1005.

Doll, R. and A.B. Hill. 1950. Smoking and carcinoma of the lung; preliminary report. *Br Med J* 2: 739-748.

Doll, R. and R. Peto. 1981. The causes of cancer: quantitative estimates of avoidable risks of cancer in the United States today. *J Natl Cancer Inst* 66: 1191-1308.

Fearon, E.R. and B. Vogelstein. 1990. A genetic model for colorectal tumorigenesis. *Cell* 61: 759-767.

Friedewald, W.F. and P. Rous. 1944. The initiating and promoting elements in tumor production : an analysis of the effects of tar, benzpyrene, and methylcholanthrene on rabbit skin. *J Exp Med* 80: 101-126.

Greenman, C. et al. 2007. Patterns of somatic mutation in human cancer genomes. *Nature* 446: 153-158.

Grollman, A.P. and M. Moriya. 1993. Mutagenesis by 8-oxoguanine: an enemy within. *Trends Genet* 9: 246-249.

Haber, D.A. and J. Settleman. 2007. Cancer: drivers and passengers. *Nature* 446: 145-146.

Harris, C.C., H. Autrup, R. Connor, L.A. Barrett, E.M. McDowell and B.F. Trump. 1976. Interindividual variation in binding of benzo[a]pyrene to DNA in cultured human bronchi. *Science* 194: 1067-1069.

Harris, C.C., V.M. Genta, A.L. Frank, D.G. Kaufman, L.A. Barrett, E.M. McDowell and B.F. Trump. 1974. Carcinogenic polynuclear hydrocarbons bind to macromolecules in cultured human bronchi. *Nature* 252: 68-69.

Hecht, S.S., C.B. Chen, N. Hirota, R.M. Ornaf, T.C. Tso and D. Hoffmann. 1978. Tobacco-specific nitrosamines: formation from nicotine in vitro and during tobacco curing and carcinogenicity in strain A mice. *J Natl Cancer Inst* 60: 819-824.

Hill, J. 1761. Cautions Against the Immoderate Use of Snuff. Founded on the Known Qualities of the Tobacco Plant; And the Effects it Must Produce when this Way Taken into the Body: And Enforced by Instances of Persons who have Perished Miserably of Diseases, Occasioned, or Rendered Incurable by its Use. In (ed. pp. R. Baldwin and J. Jackson, London.

Hollstein, M., D. Sidransky, B. Vogelstein and C.C. Harris. 1991. p53 mutations in human cancers. *Science* 253: 49-53.

Hsu, I.C., R.A. Metcalf, T. Sun, J.A. Welsh, N.J. Wang and C.C. Harris. 1991. Mutational hotspot in the p53 gene in human hepatocellular carcinomas. *Nature* 350: 427-428.

Hueper, W.C., F.H. Wiley and H.D. Wolfe. 1938. Experimental production of bladder tumors in dogs by administration of h-napthylamine. *J Ind Hyg Toxicol* 20: 46-48.

Kennaway, E.L. 1930. Further experiments on cancer-producing substances. *Biochem J* 24: 497-504.

Kennaway, E.L. and I. Hieger. 1930. Carcinogenic substances and their fluorescence spectra. *Br Med J* 1: 1044-1046.

Knudson, A. 1971. Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A* 68: 820-823.

Koreeda, M., P.D. Moore, P.G. Wislocki, W. Levin, H. Yagi and D.M. Jerina. 1978. Binding of benzo[a]pyrene 7,8-diol-9,10-epoxides to DNA, RNA, and protein of mouse skin occurs with high stereoselectivity. *Science* 199: 778-781.

Kress, M., E. May, R. Cassingena and P. May. 1979. Simian virus 40-transformed cells express new species of proteins precipitable by anti-simian virus 40 tumor serum. *J Virol* 31: 472-483.

Lane, D.P. and L.V. Crawford. 1979. T antigen is bound to a host protein in SV40-transformed cells. *Nature* 278: 261-263.

Leitch, A. and E.L. Kennaway. 1922. Experimental production of cancer by arsenic. *Br. Med. J.* 2: 1107-1108.

Ley, T.J. et al. 2008. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* 456: 66-72.

Lindahl, T. and B. Nyberg. 1972. Rate of depurination of native deoxyribonucleic acid. *Biochemistry* 11: 3610-3618.

Linzer, D.I. and A.J. Levine. 1979. Characterization of a 54K dalton cellular SV40 tumor antigen present in SV40-transformed cells and uninfected embryonal carcinoma cells. *Cell* 17: 43-52.

Liu, Z., K.R. Muehlbauer, H.H. Schmeiser, M. Hergenhahn, D. Belharazem and M.C. Hollstein. 2005. p53 mutations in benzo(a)pyrene-exposed human p53 knock-in murine fibroblasts correlate with p53 mutations in human lung tumors. *Cancer Res* 65: 2583-2587.

Loeb, L.A. and H.V. Gelboin. 1963. Stimulation of amino-acid incorporation by nuclear ribonucleic acid from normal and methylcholanthrene-treated rats. *Nature* 199: 809-810.

Loeb, L.A., C.F. Springgate and N. Battula. 1974. Errors in DNA replication as a basis of malignant changes. *Cancer Res* 34: 2311-2321.

Luo, J.L., W.M. Tong, J.H. Yoon, M. Hergenhahn, R. Koomagi, Q. Yang, D. Galendo, G.P. Pfeifer, Z.Q. Wang and M. Hollstein. 2001. UV-induced DNA damage and mutations in Hupki (human p53 knock-in) mice recapitulate p53 hotspot alterations in sun-exposed human skin. *Cancer Res* 61: 8158-8163.

Magee, P.N. and B. JM. 1956. The production of malignant primary hepatic tumours in the rat by feeding dimethylnitrosamine. *Br J Cancer* 10: 114-122.

Miller, E.C. and J.A. Miller. 1947. The presence and significance of bound amino azodyes in the livers of rats fed p- dimethyl-aminoazobenzene. 7: 468-480.

Mueller, G.C. and J.A. Miller. 1948. The metabolism of 4-dimethylaminoazobenzene by rat liver homogenates. *J Biol Chem* 176: 535-544.

Nagao, M., M. Honda, Y. Seino, T. Yahagi and T. Sugimura. 1977. Mutagenicities of smoke condensates and the charred surface of fish and meat. *Cancer Lett* 2: 221-226.

Nowell, P.C. 1976. The clonal evolution of tumor cell populations. *Science* **194**: 23-28.

Omura, T. and R. Sato. 1962. A new cytochrome in liver microsomes. *J Biol Chem* **237**: 1375-1376.

Paris, J.A. 1822. Pharmacologica; or the History of Medicinal Substances, with a View to Establish the Art of Prescribing and of Composing Extemporaneous Formulae upon Fixed and Scientific Principles. In (ed. pp. 206-217. F. & R. Lockwood, New York.

Perera, F.P., M.C. Poirier, S.H. Yuspa, J. Nakayama, A. Jaretzki, M.M. Curnen, D.M. Knowles and I.B. Weinstein. 1982. A pilot project in molecular cancer epidemiology: determination of benzo[a]pyrene--DNA adducts in animal and human tissues by immunoassays. *Carcinogenesis* **3**: 1405-1410.

Pleasance, E.D. et al. 2010a. A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* **463**: 191-196.

Pleasance, E.D. et al. 2010b. A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature* **463**: 184-190.

Pott, P. 1775. Cancer Scroti, In: Chirurgical observations relative to the cataract, the polypus of the nose, the cancer of the scrotum, the different kinds of ruptures, and the modification of the toes and feet. 63-68.

Puisieux, A., S. Lim, J. Groopman and M. Ozturk. 1991. Selective targeting of p53 gene mutational hotspots in human cancers by etiologically defined carcinogens. *Cancer Res* **51**: 6185-6189.

Ramazzini, B. 1700. De Morbis Artificum Diatriba (Diseases of workers).

Rehn, L. 1895. Blasengeschwülste bei Fuchsin-Arbeitern. *Arch. Klin. Chir.* **50**: 588-600.

Rigoni-Stern, D. 1842. Fatti statistici relativi alle malattic Cancerose. *Giorn. Prog. Patol. Terap* **2**: 507-517.

Setlow, J.K. and R.B. Setlow. 1961. Ultraviolet action spectra of ordered and disordered DNA. *Proc Natl Acad Sci U S A* **47**: 1619-1627.

Shah, S.P. et al. 2009. Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature* **461**: 809-813.

Shamberger, R.J., T.L. Andreone and C.E. Willis. 1974. Antioxidants and cancer. IV. Initiating activity of malonaldehyde as a carcinogen. *J Natl Cancer Inst* **53**: 1771-1773.

Shih, C. and R.A. Weinberg. 1982. Isolation of a transforming sequence from a human bladder carcinoma cell line. *Cell* **29**: 161-169.

Sims, P., P.L. Grover, A. Swaisland, K. Pal and A. Hewer. 1974. Metabolic activation of benzo(a)pyrene proceeds by a diol-epoxide. *Nature* **252**: 326-328.

Sporn, M.B., C.W. Dingman, H.L. Phelps and G.N. Wogan. 1966. Aflatoxin B1: binding to DNA in vitro and alteration of RNA metabolism in vivo. *Science* **151**: 1539-1541.

Theophrastus Bombastus von Hohenheim (Paracelsus). 1567. Von den Krankheiten so die Vernunfft Berauben.

Thiersch, K. 1875. Der Epithelialkrebs, namentlich der Haut. Eine anatomisch-klinische Untersuchung.

Tornaletti, S., D. Rozek and G.P. Pfeifer. 1993. The distribution of UV photoproducts along the human p53 gene and its relation to mutations in skin cancer. *Oncogene* **8**: 2051-2057.

Warburg, O., K.R. Posener and E. Negelein. 1924. The metabolism of the cancer cell. *Biochemiche Ztschr* **152**: 319-344.

Watson, J.D. and F.H. Crick. 1953. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**: 737-738.

Wynder, E.L. and E.A. Graham. 1950. Tobacco smoking as a possible etiologic factor in bronchiogenic carcinoma; a study of 684 proved cases. *J Am Med Assoc* **143**: 329-336.

Yamagiwa, K. and K. Ichikawa. 1918. Experimental study of the pathogenesis of carcinoma. *J Cancer Res* **3**: 1-21.

Yang, S.K., D.W. McCourt, P.P. Roller and H.V. Gelboin. 1976. Enzymatic conversion of benzo(a)pyrene leading predominantly to the diol-epoxide r-7,t-8-dihydroxy-t-9,10-oxy-7,8,9,10-tetrahydrobenzo(a)pyrene through a single enantiomer of r-7, t-8-dihydroxy-7,8-dihydrobenzo(a)pyrene. *Proc Natl Acad Sci U S A* **73**: 2594-2598.

Supplementary Table 2

Data used for the analysis	Exposure or cancer types	Sequencing strategy	Number of mutations / number of tumors / type of mutation used for the analysis	References
TP53 database	Lung	TP53 gene only	2994 /multiple / Syn and Non-Syn in the TP53 coding region	UMD TP53 database
Hupki mouse	Benzo(a)pyrene	TP53 gene only	33 / immortalized Hupki embryonic fibroblasts	(Liu et al., 2005)
Davies et al.	Lung	Kinome, 512 genes	145/ 26 tumors / Syn and Non-Syn in coding DNA only	(Davies et al., 2005)
Greenman et al;	Lung	Kinome, 512 genes	281/ 20 tumors / Syn and Non-Syn in coding DNA only	(Greenman et al., 2007)
Ding et al.	Lung	Exome, 623 CAN genes	1023/183/ Syn and Non-Syn in coding DNA only	(Ding et al., 2008)
Kan et al.	Lung	Exome, 1507 genes	1561/134/ Syn and Non-Syn in coding DNA only	(Kan et al., 2010)
Pleasance et al.	NCI-H209, SCLC cell line	Whole genome sequencing	50 274/1/ all types of substitutions in coding and non-coding DNA	(Pleasance et al., 2010b)
Lee et al.	Lung, NSCLC	Whole genome sequencing	50 675/1/all types of substitutions in coding and non-coding DNA	(Lee et al., 2010)
TP53 database	Colorectal	TP53 gene only	3894/ multiple / Syn and Non-Syn in the TP53 coding region	UMD TP53 database
Sjöblom et al.	Colorectal	Exome, all genes	729/11/ Syn and Non-Syn in coding DNA only	(Sjoblom et al., 2006)
Greenman et al.	Colorectal	Kinome, 512 genes	166/28T/ Syn and Non-Syn in coding DNA only	(Greenman et al., 2007)
TP53	Glioblastoma	TP53 gene only	757/multiple / Syn and Non-Syn in the TP53 coding region	UMD TP53 database
Parson et al.	Glioblastoma	Exome / all genes	453/22/ Syn and Non-Syn in coding DNA only	(Parsons et al., 2008)
TCGA	Glioblastoma	Exome / all genes	2355/206/ Syn and Non-Syn in coding DNA only	(McLendon R et al., 2008)

HGMD		2113 genes associated with hereditary disease	43953 / all types of substitutions in coding and non-coding DNA	(Stenson et al., 2009)
TP53 database	HCC	TP53 gene only	1091/multiple / Syn and Non-Syn in the TP53 coding region	UMD TP53 database
RIKEN (JP)	HCC	Whole genome sequencing (partial)	162/1/ all types of mutations in coding and non-coding DNA	http://dcc.icgc.org/ (version 3)
NCC (JP)	HCC	Whole genome sequencing (partial)	493/1 / all types of mutations in coding and non-coding DNA	http://dcc.icgc.org/ (version 3)
TP53 database	Melanoma	TP53 gene only	113/multiple / Syn and Non-Syn in the TP53 coding region	UMD TP53 database
Pleasance et al.	Colo-829 cell line	Whole genome sequencing	76085/1/all types of mutations in coding and non-coding DNA	(Pleasance et al., 2010a)
TP53 database	Breast	TP53 gene only	3367/ multiple/ Syn and Non-Syn in the TP53 coding region	UMD TP53 database
Stephens et al.	Breast	Kinome, 518 genes	90/ 25/ Syn and Non-Syn in coding DNA only	(Stephens et al., 2009)
Sjöblom et al.	Breast	Exome, all genes	904/ 11/ Syn and Non-Syn in coding DNA only	(Sjoblom et al., 2006)
Greenman et al.	Breast	Kinome, 512 genes	103/16/ Syn and Non-Syn in coding DNA only	(Greenman et al., 2007)
Kan et al.	Breast	Exome, 1507 CAN genes	654/183/ Syn and Non-Syn in coding DNA only	(Kan et al., 2010)
Shah et al.	Breast	Whole genome sequencing	1328/1/ all types of mutations in coding and non-coding DNA	(Shah et al., 2009)

REFERENCES

Davies, H. et al. 2005. Somatic mutations of the protein kinase gene family in human lung cancer. *Cancer Res* **65**: 7591-7595.

Ding, L. et al. 2008. Somatic mutations affect key pathways in lung adenocarcinoma. *Nature* **455**: 1069-1075.

Greenman, C. et al. 2007. Patterns of somatic mutation in human cancer genomes. *Nature* **446**: 153-158.

Kan, Z. et al. 2010. Diverse somatic mutation patterns and pathway alterations in human cancers. *Nature* 466: 869-873.

Lee, W. et al. 2010. The mutation spectrum revealed by paired genome sequences from a lung cancer patient. *Nature* 465: 473-477.

Liu, Z., K.R. Muehlbauer, H.H. Schmeiser, M. Hergenhahn, D. Belharazem and M.C. Hollstein. 2005. p53 mutations in benzo(a)pyrene-exposed human p53 knock-in murine fibroblasts correlate with p53 mutations in human lung tumors. *Cancer Res* 65: 2583-2587.

McLendon R, et al. 2008. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* 455: 1061-1068.

Parsons, D.W. et al. 2008. An integrated genomic analysis of human glioblastoma multiforme. *Science* 321: 1807-1812.

Pleasance, E.D. et al. 2010a. A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* 463: 191-196.

Pleasance, E.D. et al. 2010b. A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature* 463: 184-190.

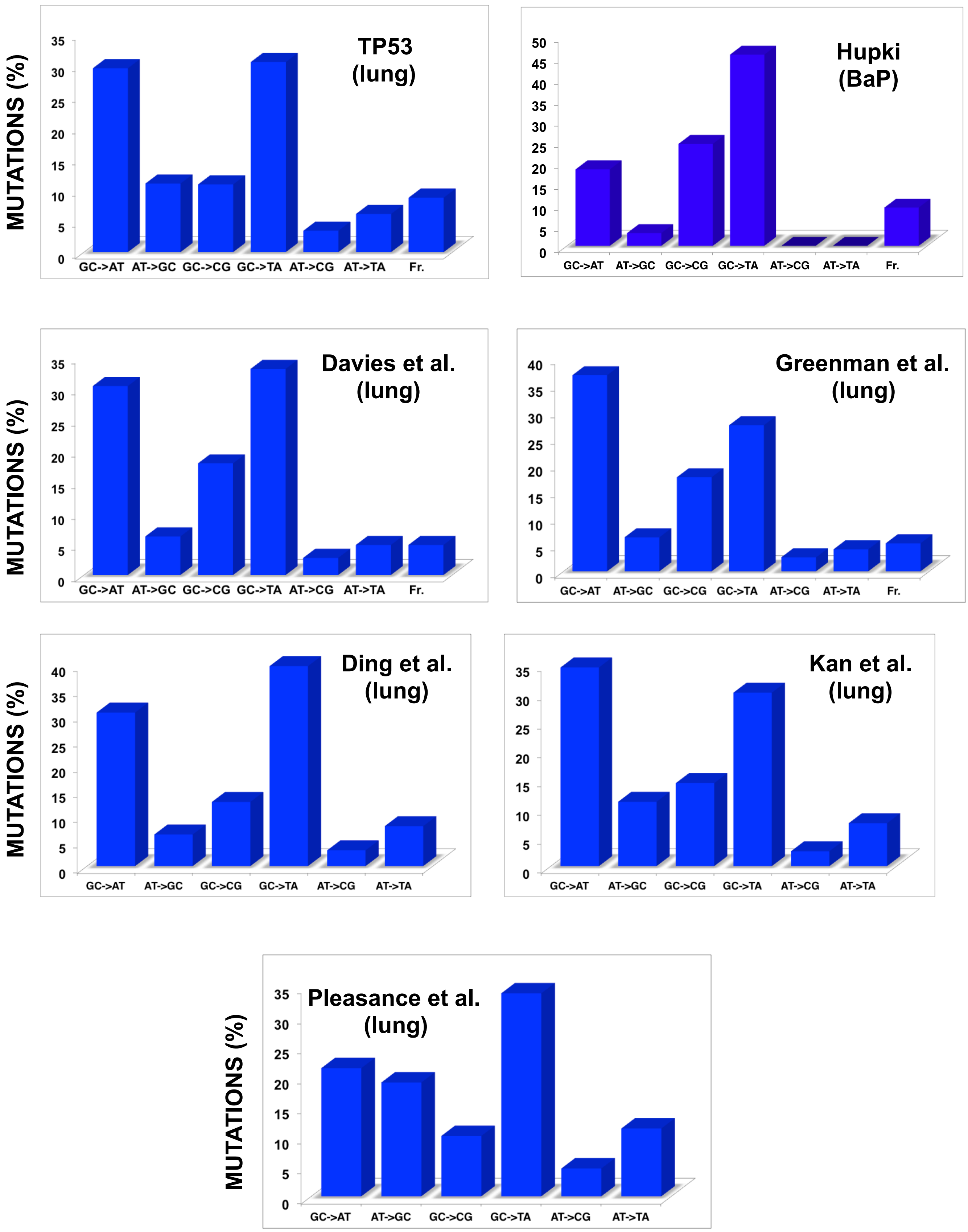
Shah, S.P. et al. 2009. Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature* 461: 809-813.

Sjoblom, T. et al. 2006. The consensus coding sequences of human breast and colorectal cancers. *Science* 314: 268-274.

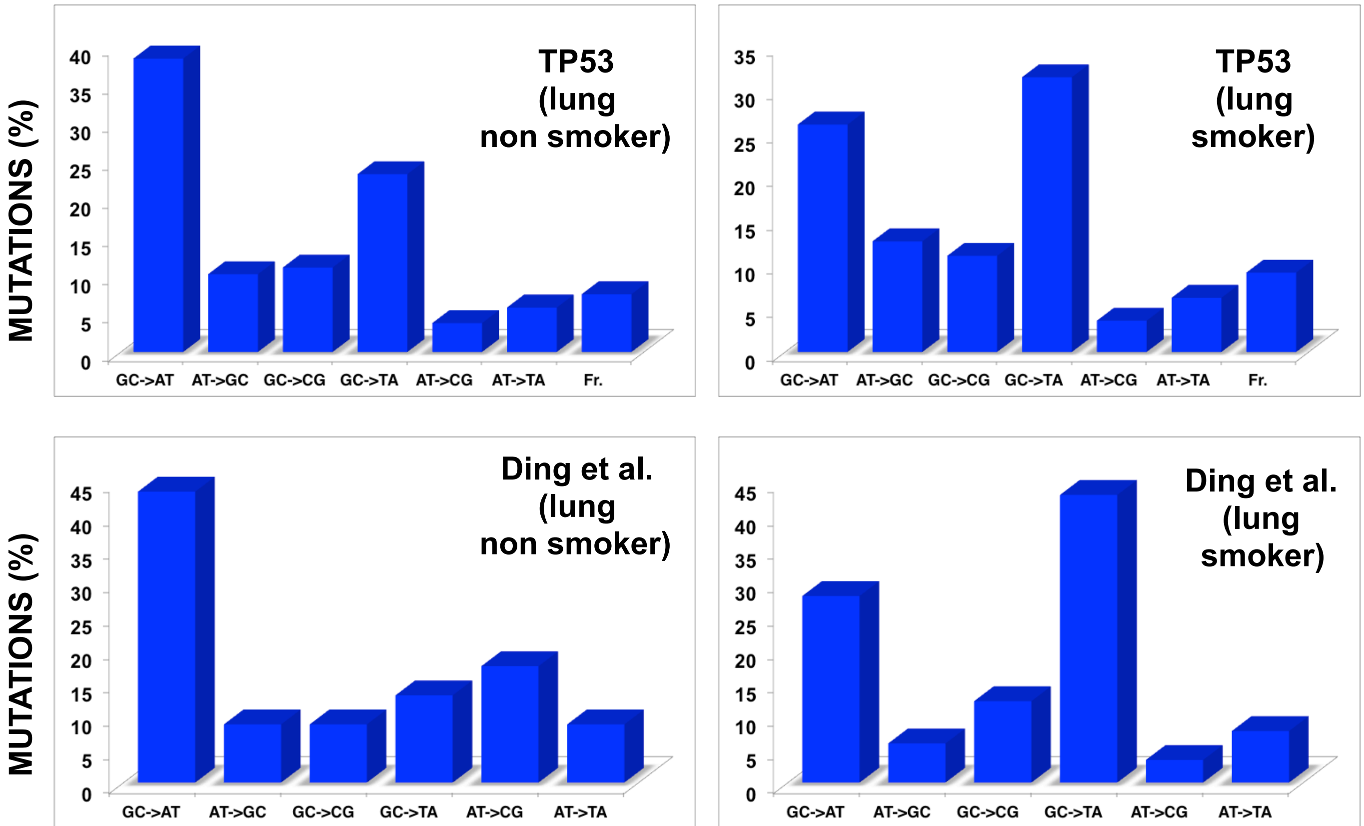
Stenson, P.D., E.V. Ball, K. Howells, A.D. Phillips, M. Mort and D.N. Cooper. 2009. The Human Gene Mutation Database: providing a comprehensive central mutation database for molecular diagnostics and personalized genomics. *Hum Genomics* 4: 69-72.

Stephens, P.J. et al. 2009. Complex landscapes of somatic rearrangement in human breast cancer genomes. *Nature* 462: 1005-1010.

Sup Fig 1A

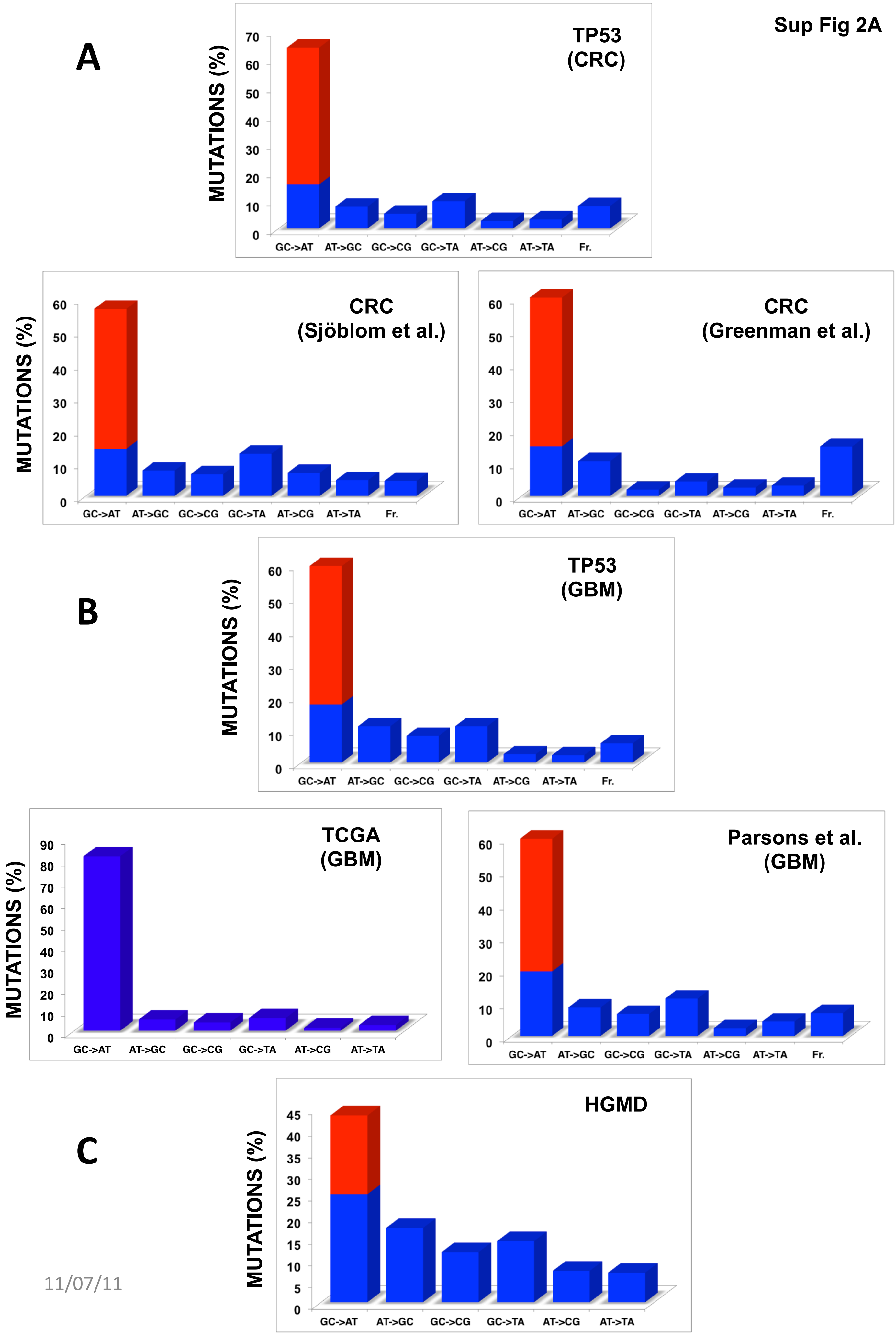

MUTATIONS (%)
TP53 (lung)
Hupki (BaP)
Davies et al. (lung)
Greenman et al. (lung)
Ding et al. (lung)
Kan et al. (lung)
Pleasance et al. (lung)
GC->AT
AT->GC
GC->CG
GC->TA
AT->CG
AT->TA
Fr.

Sup Fig 1B



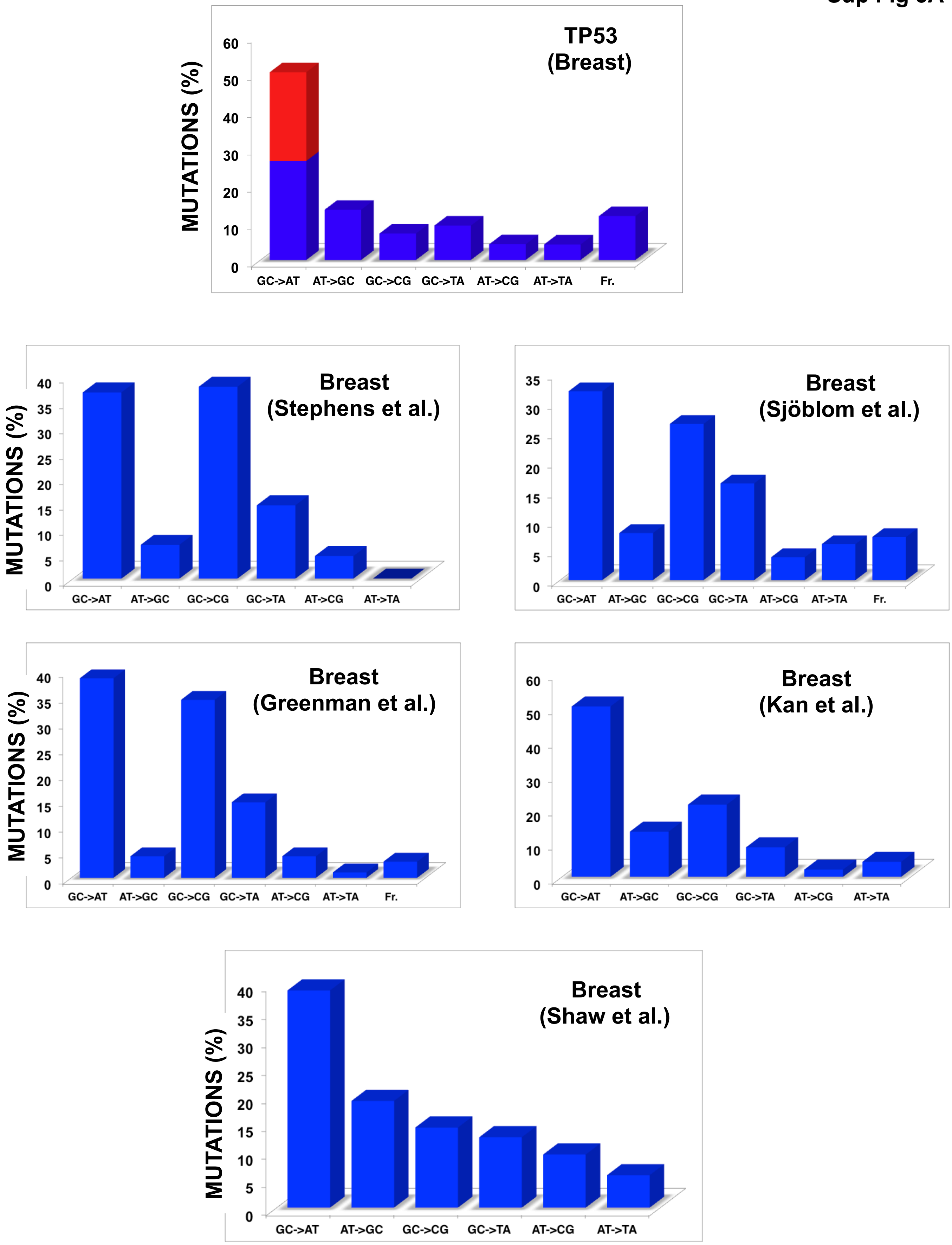
MUTATIONS (%)
TP53 (lung non smoker)
0 5 10 15 20 25 30 35 40
GC->AT AT->GC GC->CG GC->TA AT->CG AT->TA Fr.
TP53 (lung smoker)
0 5 10 15 20 25 30 35
GC->AT AT->GC GC->CG GC->TA AT->CG AT->TA Fr.
MUTATIONS (%)
Ding et al. (lung non smoker)
0 5 10 15 20 25 30 35 40 45
GC->AT AT->GC GC->CG GC->TA AT->CG AT->TA
Ding et al. (lung smoker)
0 5 10 15 20 25 30 35 40 45
GC->AT AT->GC GC->CG GC->TA AT->CG AT->TA

Sup Fig 2A

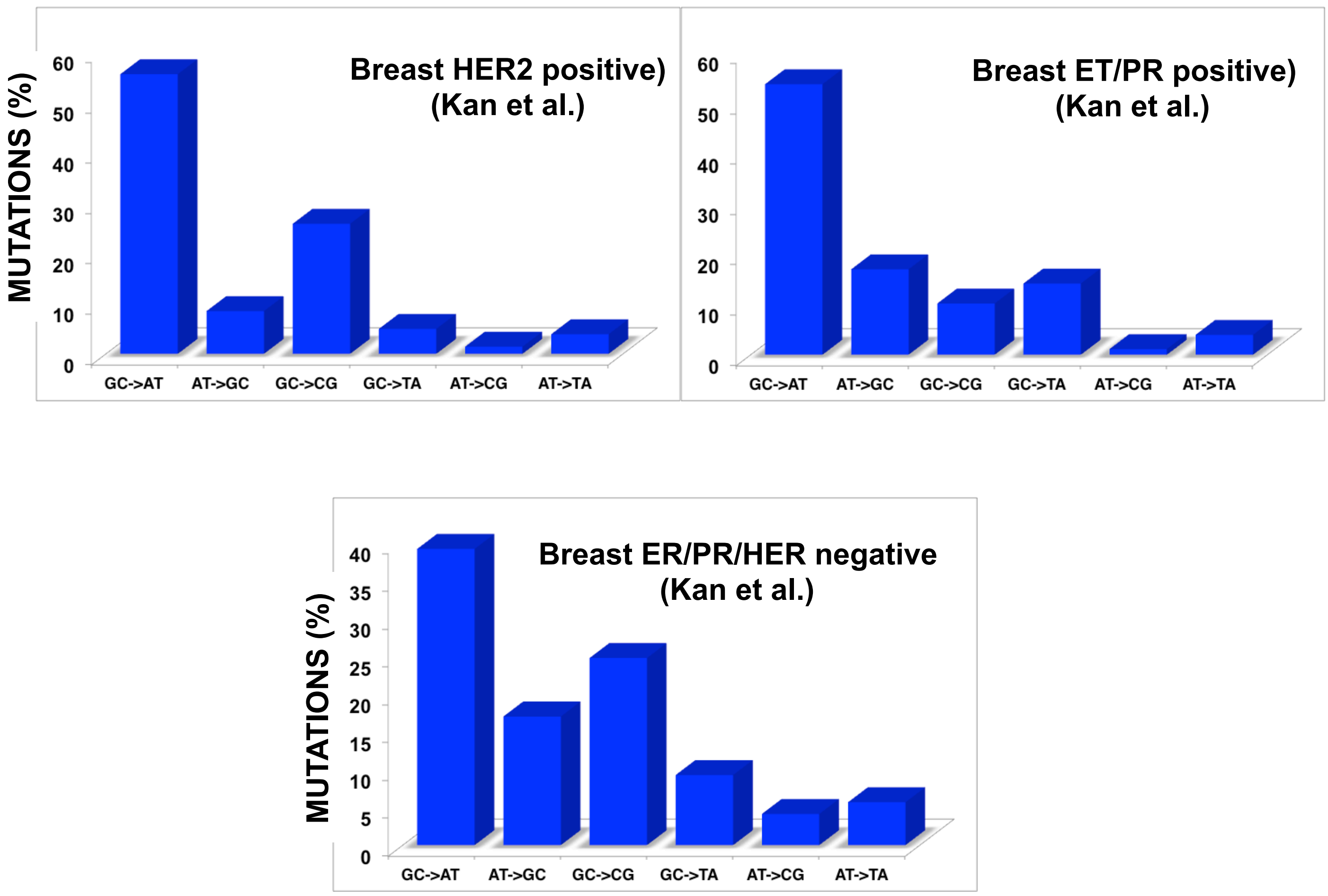


11/07/11

Sup Fig 3A

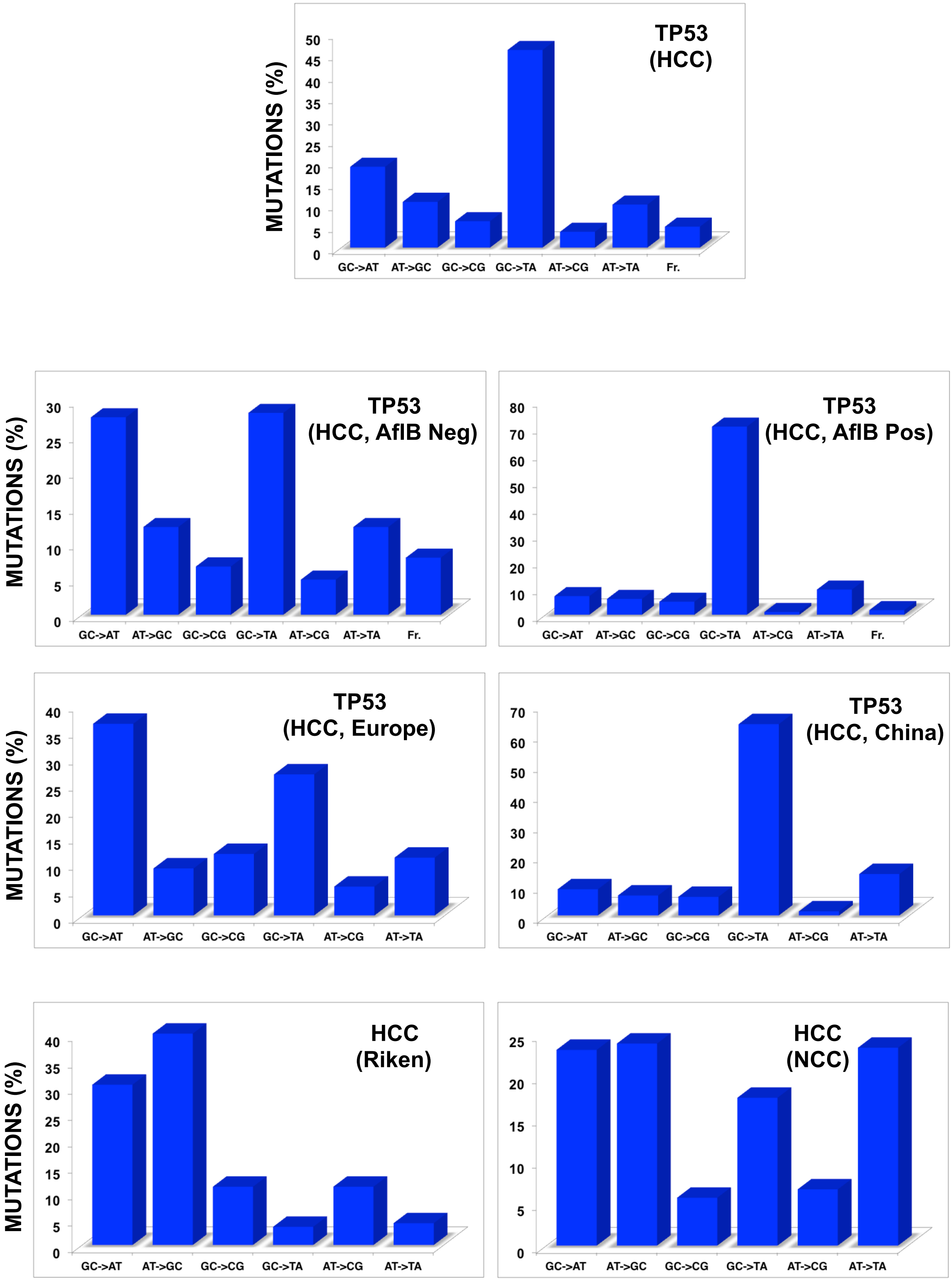

TP53
(Breast)
MUTATIONS (%)
0
10
20
30
40
50
60
GC->AT
AT->GC
GC->CG
GC->TA
AT->CG
AT->TA
Fr.
Breast
(Stephens et al.)
MUTATIONS (%)
0
5
10
15
20
25
30
35
40
GC->AT
AT->GC
GC->CG
GC->TA
AT->CG
AT->TA
Breast
(Sjöblom et al.)
0
5
10
15
20
25
30
35
GC->AT
AT->GC
GC->CG
GC->TA
AT->CG
AT->TA
Fr.
Breast
(Greenman et al.)
MUTATIONS (%)
0
5
10
15
20
25
30
35
40
GC->AT
AT->GC
GC->CG
GC->TA
AT->CG
AT->TA
Fr.
Breast
(Kan et al.)
0
10
20
30
40
50
60
GC->AT
AT->GC
GC->CG
GC->TA
AT->CG
AT->TA
Breast
(Shaw et al.)
MUTATIONS (%)
0
5
10
15
20
25
30
35
40
GC->AT
AT->GC
GC->CG
GC->TA
AT->CG
AT->TA

Sup Fig 3B



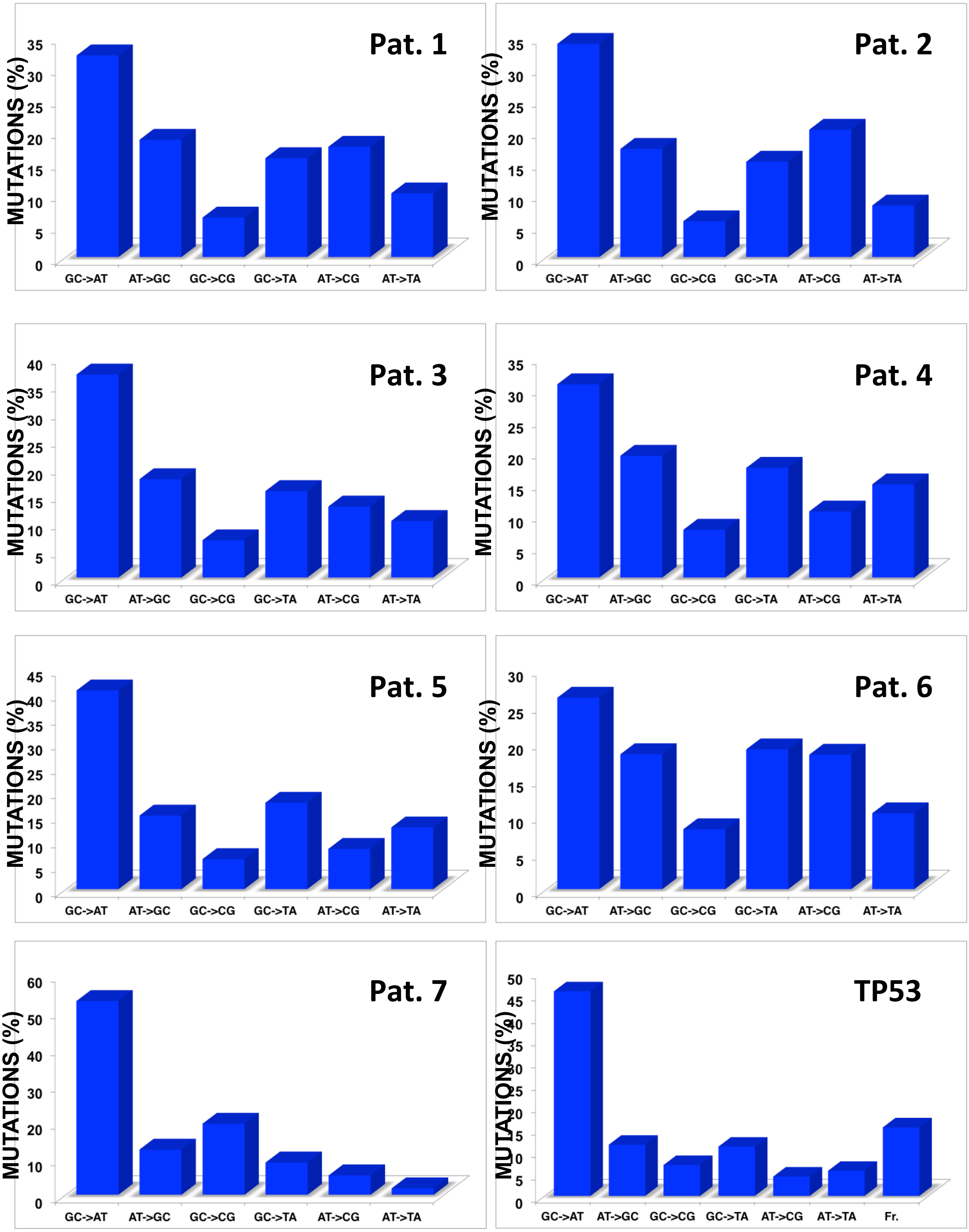
11/07/11

Sup Fig 4A

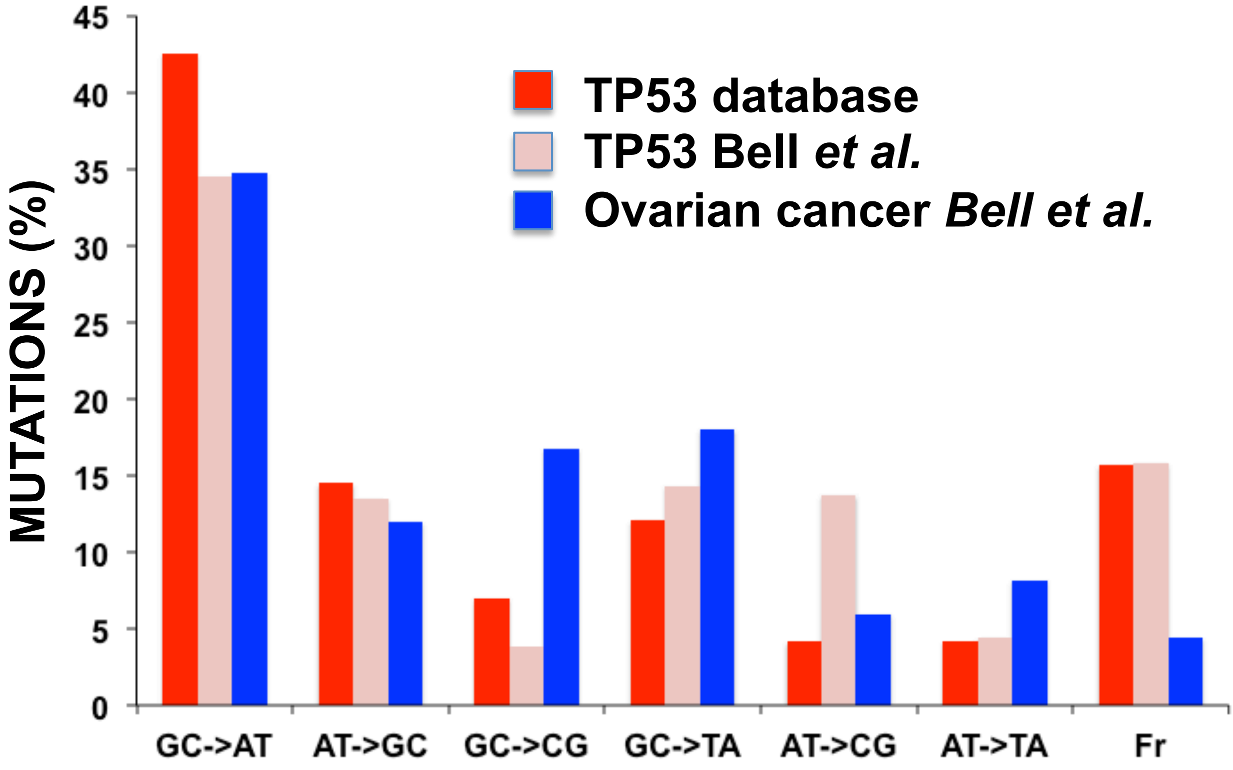


11/07/11

Sup Fig 5



11/07/11



45
40
35
30
25
20
15
10
5
0
MUTATIONS (%)
TP53 database
TP53 Bell *et al.*
Ovarian cancer *Bell et al.*
GC->AT
AT->GC
GC->CG
GC->TA
AT->CG
AT->TA
Fr

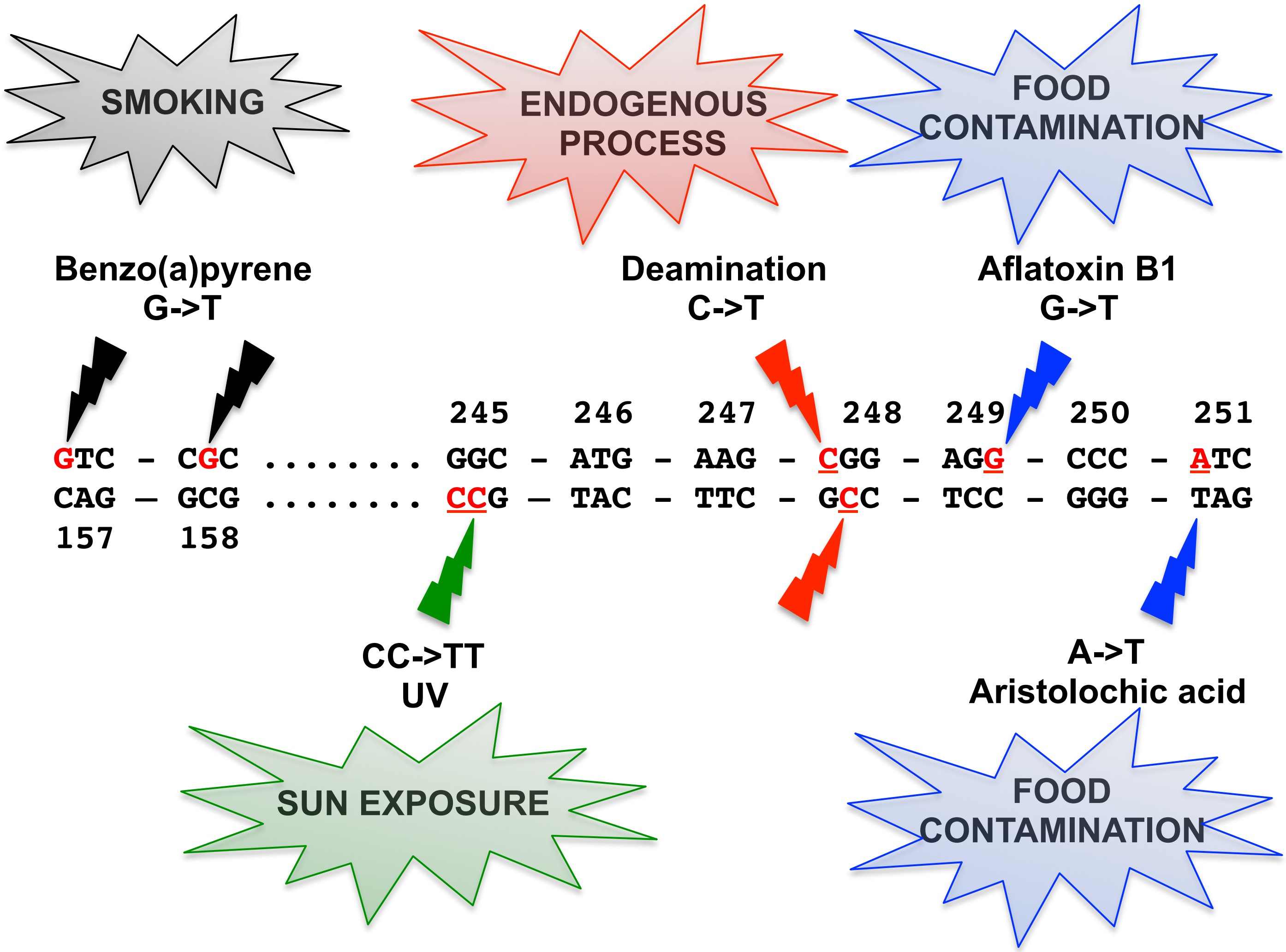
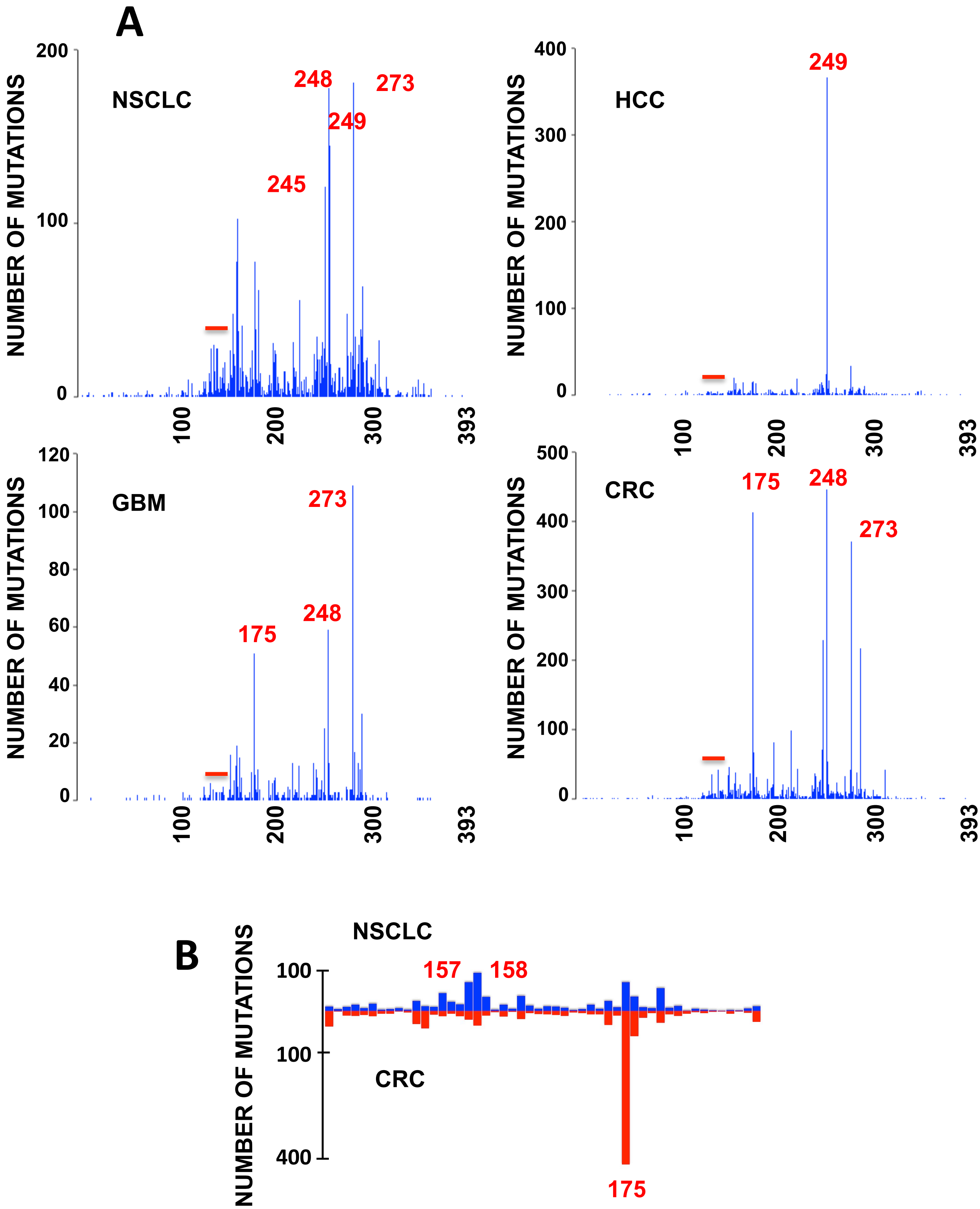

SMOKING
ENDOGENOUS PROCESS
FOOD CONTAMINATION
Benzo(a)pyrene
G->T
Deamination
C->T
Aflatoxin B1
G->T
245 246 247 248 249 250 251
GTC – CGC .... GGC – ATG – AAG – CGG – AGG – CCC – ATC
CAG – GCG .... CCG – TAC – TTC – GCC – TCC – GGG – TAG
157 158
CC->TT
UV
SUN EXPOSURE
A->T
Aristolochic acid
FOOD CONTAMINATION

Figure 2


A
NSCLC
HCC
GBM
CRC
NUMBER OF MUTATIONS
200
100
0
400
300
500
120
80
60
40
20
248
273
249
245
175
100
200
300
393
B
157
158
400

11/07/11

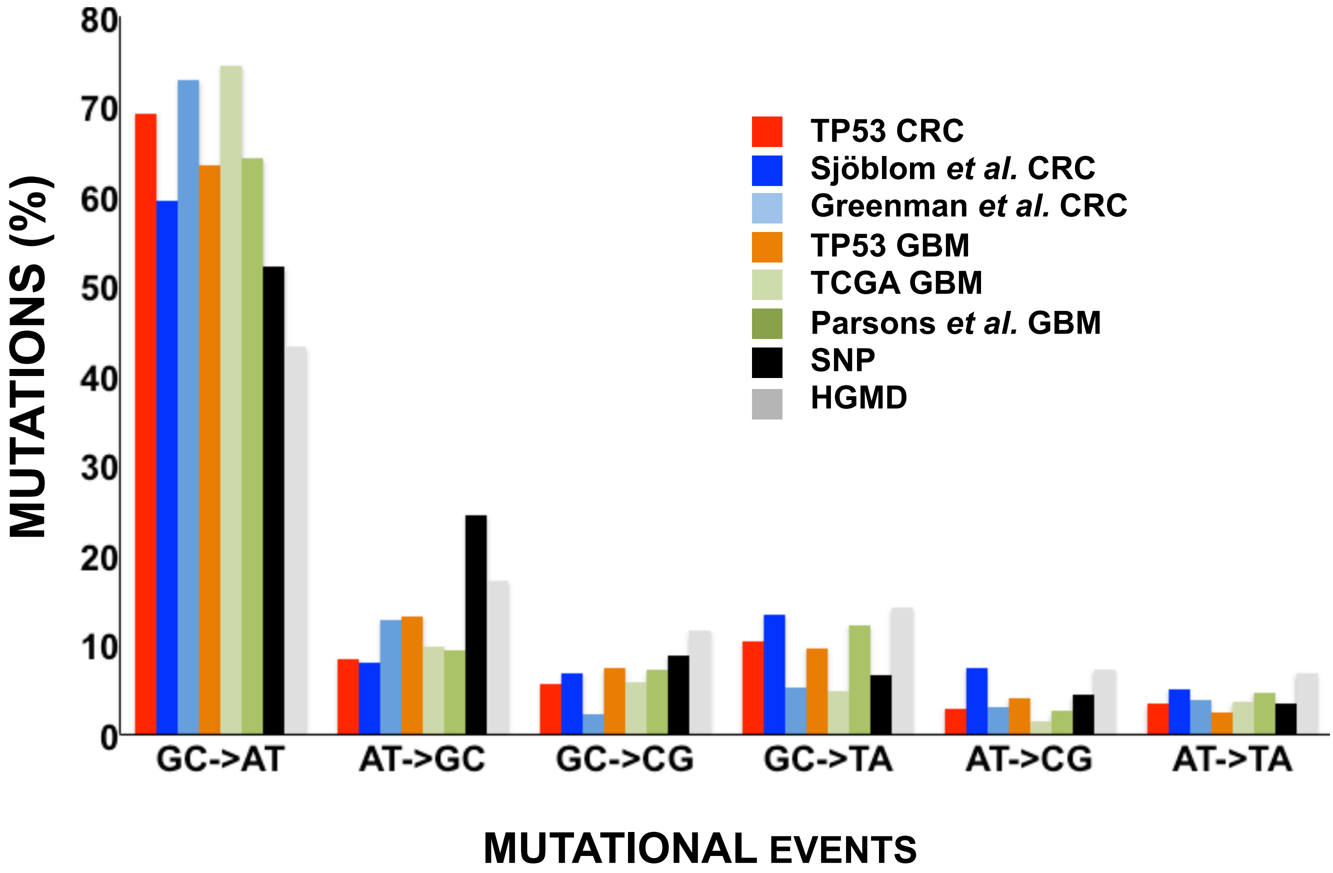
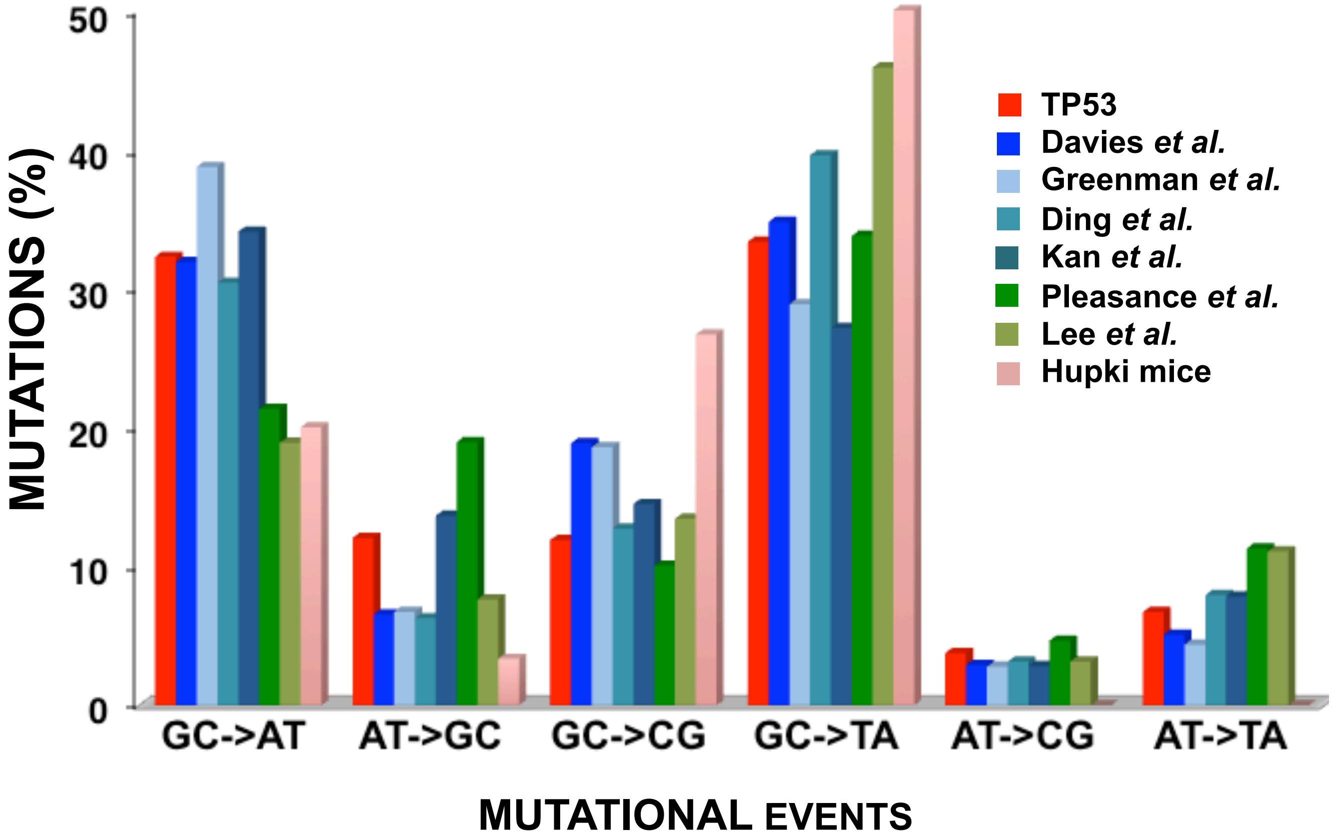

80
70
60
50
40
30
20
10
0
MUTATIONS (%)
TP53 CRC
Sjöblom *et al.* CRC
Greenman *et al.* CRC
TP53 GBM
TCGA GBM
Parsons *et al.* GBM
SNP
HGMD
GC->AT
AT->GC
GC->CG
GC->TA
AT->CG
AT->TA
MUTATIONAL EVENTS

Figure 4

Sup Fig 10

A



B

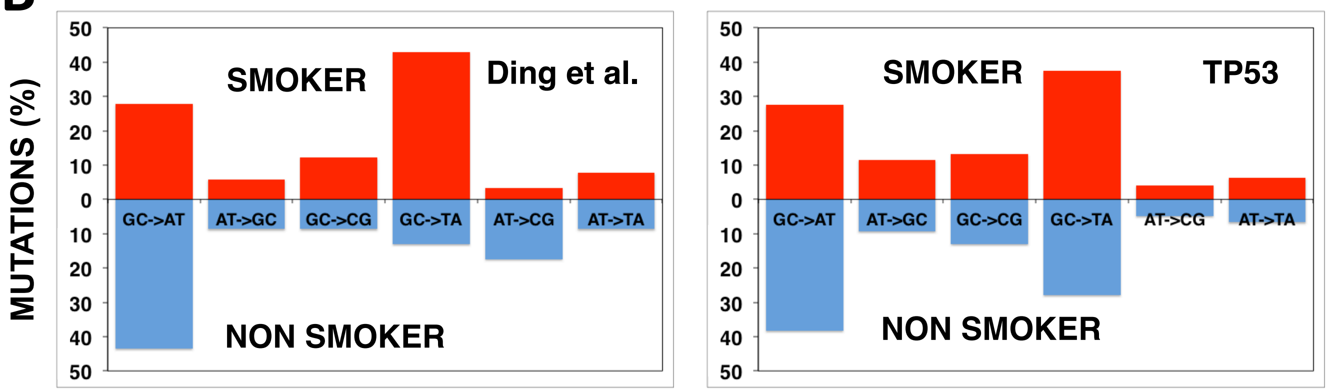
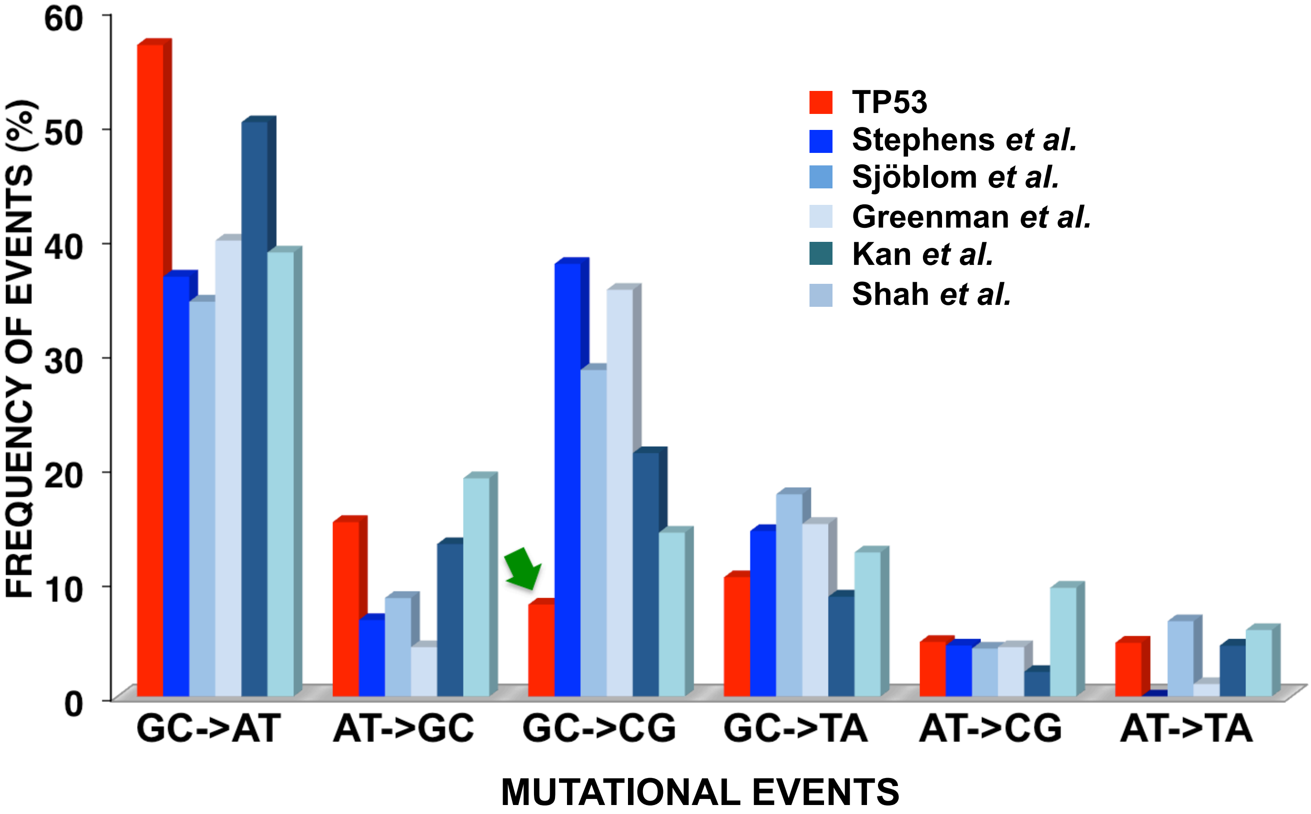
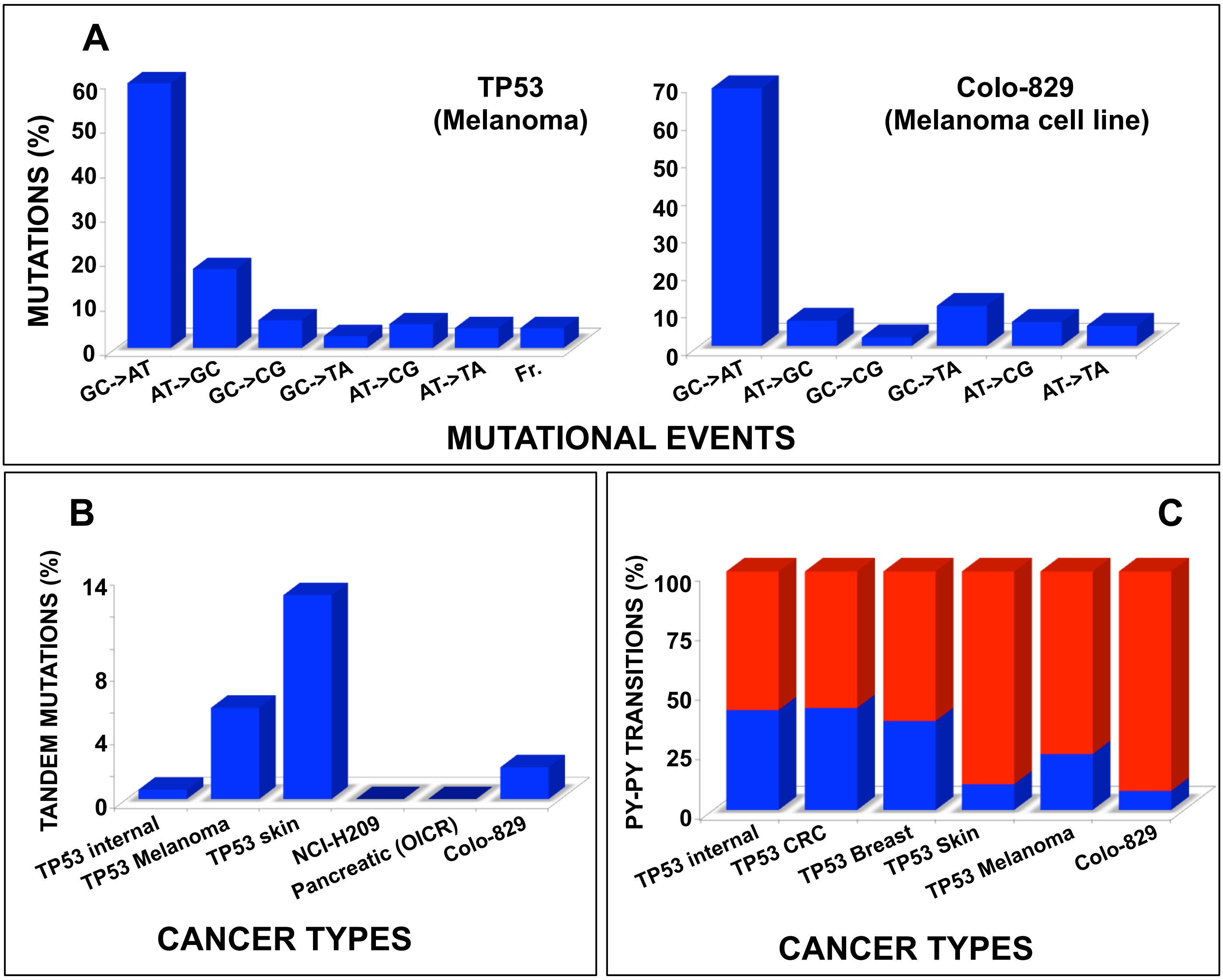


Figure 5



11/07/11

Figure 6



11/07/11

Figure 7

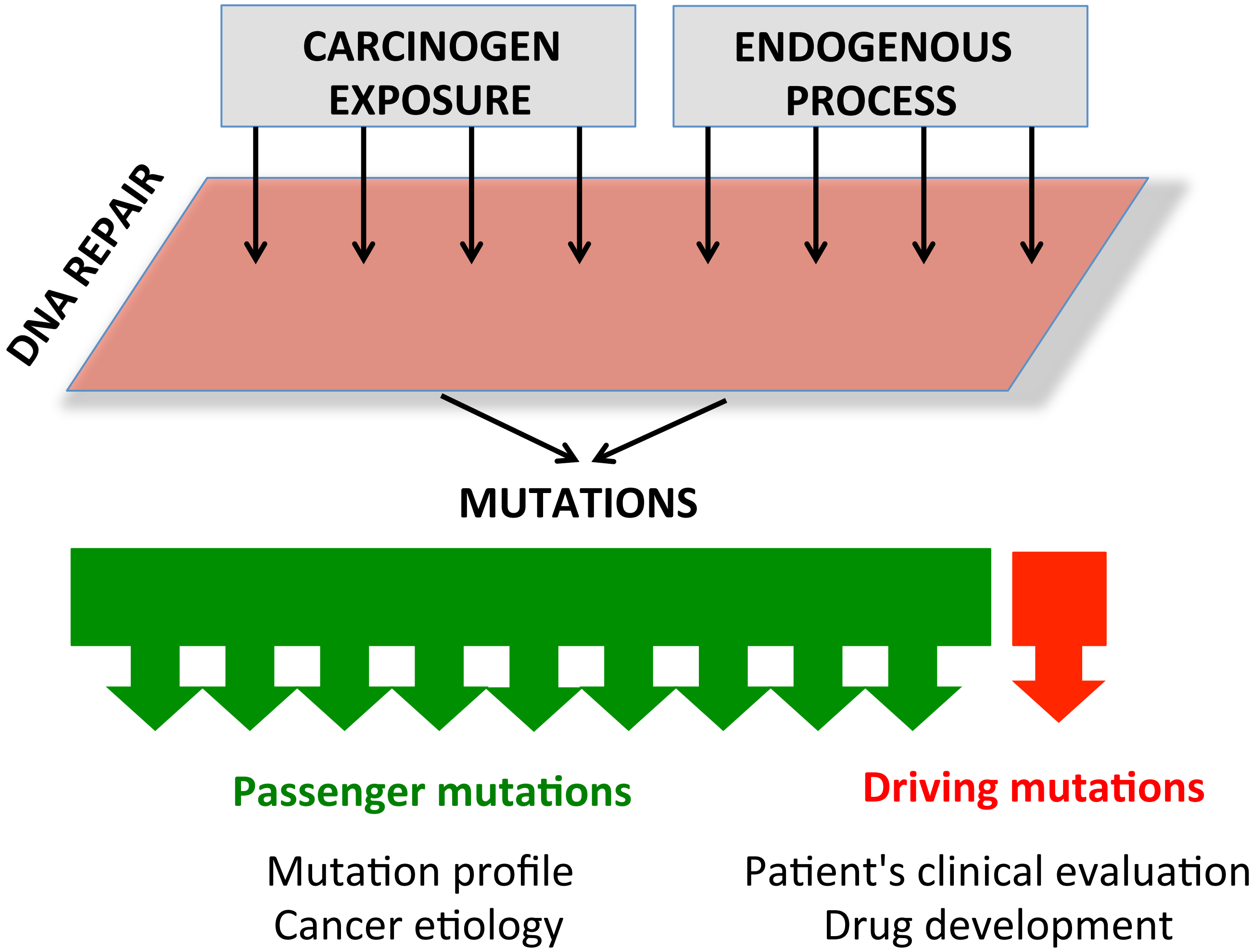

CARCINOGEN
EXPOSURE
ENDOGENOUS
PROCESS
DNA REPAIR
MUTATIONS
Passenger mutations
Driving mutations
Mutation profile
Cancer etiology
Patient's clinical evaluation
Drug development

Figure 8