3 DETECTION AND ANALYSIS OF SEQUENCE VARIANTS IN THE NCI-60 CELL LINE SET

3.1 Introduction

The NCI-60 cell lines are composed of 60 human cancer cell lines representing nine tissues-of-origin. The National Cancer Institute's Developmental Therapeutics Program (DTP) assembled the NCI-60 cell lines for an *in vitro* anti-cancer drug screen. More than 100,000 compounds have been tested for anti-cancer activity in the NCI-60. Measures of the concentration required for 50% inhibition of growth (GI_{50}) of approximately 42,000 compounds tested in the NCI-60 have been stored in a public database.

Recent studies have indicated that mutations in cancer genes may be determinants of sensitivity to targeted therapeutics against the mutated gene product. Examples include the *BCR-ABL* fusion gene in chronic myeloid leukemia (CML) that predicts sensitivity to imatinib and kinase domain mutations of *EGFR* in lung cancer that predict sensitivity to gefitinib and erlotinib. Moreover, there is evidence that mutations in cancer genes can indirectly affect the response to chemotherapy. For example, resistance to Herceptin, a monoclonal antibody against over expressed or amplified ERBB2 protein, is associated with the presence of mutations of *PTEN*, a downstream signaling effector of ERBB2.

The Cancer Gene Census reports there are approximately 363 cancer genes (http://www.sanger.ac.uk/genetics/CGP/Census/). Prior to the studies in this thesis, the NCI-60 cell lines had only been analyzed for mutations in *HRAS*, *KRAS*, *NRAS* and *TP53*. In order to extend this analysis and hence to allow further exploration of the interaction between drug activity and mutations in cancer genes, I undertook sequencing of a much larger selection of cancer genes. This chapter reports on the results of sequencing 24 cancer genes in the NCI-60 cell lines.

3.2 Classification of sequence variants

There are no matched normal DNA samples for the NCI-60 cell lines with which to determine the provenance of the observed variants. In particular, it is impossible to determine directly whether a variant was somatically acquired or was present in the germline. Even if a mutation is known to be somatic, however, the presence of passenger somatic mutations in cancer genomes can render problematic understanding of the functional relevance of a particular variant. I therefore developed a pragmatic classification scheme for variants found in cancer cell lines based on information from previous mutational screens of both cancers and normal tissues. The classification has four strata:

i) Likely oncogenic mutation (LOM): a sequence variant that had previously been shown to be a somatic mutation in human cancer or was consistent with the position and type of mutations for a given gene. This class also included homozygous deletions in tumor suppressor genes;

ii) Tentative oncogenic variant (TOV): a sequence variant which, though of similar type and location to known cancer causing mutations, is different from those previously reported. This category also included heterozygous truncating variants in tumor suppressor genes but did not include heterozygous missense variants in TP53, which were classified as LOMs;

iii) Variant of unknown significance (VUS): a sequence change that was not previously reported as a single nucleotide polymorphism (SNP) and does not fit the criteria for LOMs or TOVs;

iv) Single nucleotide polymorphism (SNP): a sequence change previously reported as a germline SNP in the SNP database (dbSNP), the published literature or in our own studies.

3.3 Results of the mutation analysis of 24 cancer genes in the NCI-60 cell lines

The results of this thesis chapter concentrate on the LOMs and TOVs I identified in the NCI-60. I verified the presence of the LOMs and TOVs in the NCI-60 in a second independent amplification and sequencing experiment. However, the observed variants of unknown significance (VUS) and SNPs did not go through a second validation experiment because of the numbers of these and constraints of time.

3.3.1 APC

A total of nine *APC* variants were identified in seven of the NCI-60 cell lines (Table 3-1). Six of the seven cell lines are colon cancer derived and one

is an ovarian cancer cell line. A designation of LOM or TOV was attributed to the variants identified based on the afore-mentioned criteria. The variants of *APC* in the NCI-60 are characterized by nonsense and frame-shift mutations resulting in a truncated protein product. Homozygous variants of *APC* are designated as LOM, as are two heterozygous truncating variants co-occurring in the same cell line. Single heterozygous variants, however, are classified as TOV.

Cell line	Tumor type	*APC* variants identified	Type
COLO-205	Colorectal carcinoma	Hom c.4666_4667insA p.T1556fsX3	LOM
HCC2998	Colorectal carcinoma	Het c.1994T>A p.L665X, Het c.4348C>T R1450X	LOM
HCT-15	Colorectal carcinoma	Het c.6496C>T p.R2166X, Hom c.4248delC p.I1417fsX2	LOM
HT-29	Colorectal carcinoma	Het c.2557G>T p.E853X, Het c.4666_4667insA p.T1556fsX3	LOM
KM12	Colorectal carcinoma	Het c.5454_5455insA p.N1819fsX7	TOV
SK-OV-3	Ovarian carcinoma	Het c.4666delA p.T1556fsX9	TOV
SW620	Colorectal carcinoma	Hom c.4012C>T p.Q1338X	LOM

Table 3-1. Mutations/variants of APC identified in the NCI-60 cell lines. Legend: Hom- homozygous; Het- heterozygous; del- deletion; ins- insertion; LOM- likely oncogenic mutation; TOV- tentative oncogenic variant. This same legend applies to all other genes sequenced in the NCI-60.

The classification of TOV for *APC* variants identified in SKOV3 and KM12 cell lines was applied because of the presence of heterozygous truncating variants. Because *APC* is a tumor suppressor gene, following the Knudson two- hit model, both alleles of a tumor suppressor gene should be inactivated to cause tumorigenesis. It is possible that the second allele of *APC* in these two lines is inactivated by other mechanisms such as promoter

methylation that I have not addressed. It is also possible that there is another heterozygous truncating variant of the second allele I have not detected, as all exons were not fully amenable to sequencing. Considering that KM12 is a colorectal cancer cell line, the presence of *APC* mutation may be likely oncogenic. It is rare, however, for mutations of *APC* to be found in ovarian cancers such as SK-OV-3.

***APC* cDNA sequence:**

A. *APC* SK-OV-3 deletion of nucleotide 4666:

aatgaaaaccaagagaaagaggcagaaaaaactattgattctgaa
N E N Q E K E A E K T I D S E
aaggacctat**tag**atgattcagatgatgatgatattgaaatacta
K D L L D D S D D D D I E I L

B. *APC* KM12 insertion between nucleotides 5454 and 5455:

ttgaaaaat**^a**aattccaaggtcttcaat**tga**taagctcccaaataat
L K N N S K V F N D K L P N N

Figure 3-1. Illustration of nucleotide sequence changes in APC of cell lines with TOVs, SK-OV-3 and KM12, in the NCI-60. A) SKOV3 a microsatellite unstable (MSI+) ovarian cancer cell line harbors a deletion of an adenine at nucleotide position 4666 (bold red). This deletion occurs in a repeat sequence of six adenines (grey highlight) and causes a frame-shift resulting in a truncated protein product (black bold). B) KM12 a microsatellite unstable (MSI+) colorectal cancer cell line harbors an insertion of an adenine between nucleotide positions 5454 and 5455 (bold red) within a polynucleotide

sequence of adenine. This insertion causes a frame-shift resulting in a truncated protein product (bold black).

Alternatively, the presence of the heterozygous truncating variants of *APC* may not be oncogenic at all and may be due to the microsatellite instability of SKOV3 and KM12 (Table 3-2). The frame-shift mutation of *APC* in SK-OV-3 occurs in a stretch of six adenines, typical of the genomic regions in which DNA replication mistakes occur in MSI+ genomes (Figure 3-1). These mistakes can result in truncation of a gene product. Although KM12 is a colorectal cancer cell line in which *APC* mutations are common, the presence of microsatellite instability and the heterozygosity of the truncating variant casts uncertainty on the role of the *APC* variant in KM12. Also, the presence of *APC* variants in ovarian cancers is rare. The presence of microsatellite instability suggests that the truncating *APC* variant in SK-OV-3 may not be oncogenic and therefore justifies the TOV status.

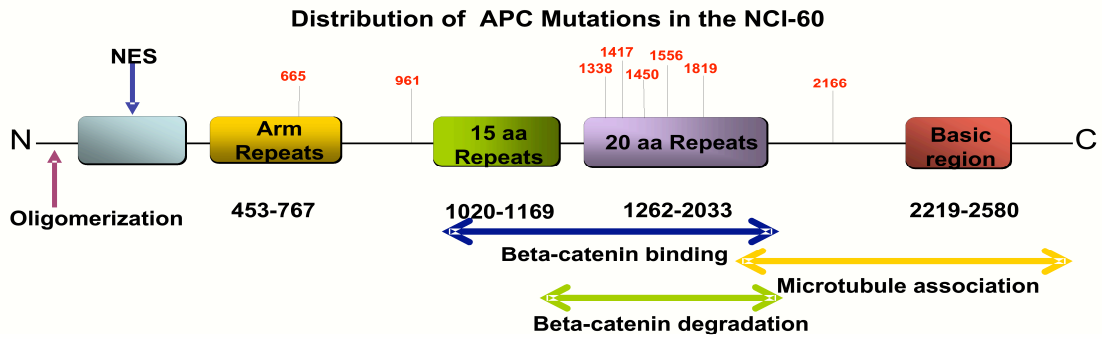

Figure 3-2. Distribution of APC mutations in the NCI-60 cell lines. A schematic of the APC protein, domains and location of amino acid positions of the truncating mutations identified in NCI-60 cell lines.

The pattern and distribution of *APC* mutations identified in the NCI-60 is similar to that reported in primary tumors (COSMIC), (Figure 3-2). The mutations of *APC* are mostly homozygous and truncating and occur most frequently in colorectal cancers. The mutations of *APC* affect the beta-catenin binding and degradation region (mutation cluster region) (COSMIC). Codon 1450 of the *APC* reading frame seems to be the most commonly affected by nonsense mutations in colorectal cancer (COSMIC) and one of the seven colorectal cancer cell lines of the NCI-60, HCC2998, harbors a nonsense mutation at codon 1450.

Microsatellite Instability Status of the NCI-60 Cell lines		
Cell line	**Tissue type**	**MSI**
786-0	Renal cell carcinoma	MSS
A498	Renal cell carcinoma	MSS
A549	Lung carcinoma	MSS
ACHN	Renal cell carcinoma	MSS
BT-549	Breast carcinoma	MSS
CAKI-1	Renal cell carcinoma	MSS
CCRF-CEM	Acute lymphoblastic leukemia	MSI-H
COLO-205	Colorectal cancer	MSS
DU-145	Prostate carcinoma	MSI-H
EKVX	Lung carcinoma	MSS
HCC2998	Colorectal carcinoma	MSS
HCT-116	Colorectal carcinoma	MSI-H
HCT-15	Colorectal carcinoma	MSI-H
HL-60	Acute myeloid leukemia	MSS
HOP62	Lung carcinoma	MSS
HOP-92	Lung carcinoma	MSS
Hs-578-T	Breast carcinoma	MSS
HT-29	Colorectal carcinoma	MSS
IGROV-1	Ovarian carcinoma	MSI-H
K-562	Chronic myeloid leukemia	MSS
KM12	Colorectal carcinoma	MSI-H
LOXIMVI	Melanoma	NA
M14/MDA-MB-435	Melanoma	MSS
MALME-3M	Melanoma	MSS
MCF7	Breast carcinoma	MSS
MDA-MB-231	Breast carcinoma	MSS
MOLT-4	Acute lymphoblastic leukemia	MSI-H

NCI-H226	Lung squamous cell carcinoma	MSS
NCI-H23	Lung adenocarcinoma	MSS
NCI-H322M	Lung bronchoalveolar carcinoma	MSS
NCI-H460	Lung large cell carcinoma	MSS
NCI-H522	Lung adenocarcinoma	MSS
OVCAR3	Ovarian carcinoma	MSS
OVCAR-4	Ovarian carcinoma	MSS
OVCAR-5	Ovarian carcinoma	MSS
OVCAR-8/ NCIADR-RES	Ovarian carcinoma	MSS
PC-3	Prostate carcinoma	MSS
RPMI-8226	Myeloma	MSS
RXF393	Renal cell carcinoma	MSS
SF-268	Glioma	MSS
SF-295	Glioma	MSS
SF539	Glioma	MSS
SK-MEL-2	Melanoma	MSI-H
SK-MEL-28	Melanoma	MSS
SK-MEL-5	Melanoma	MSS
SK-OV-3	Ovarian carcinoma	MSI-L
SN12C	Renal cell carcinoma	MSS
SNB-75	Glioma	MSS
SR	Non Hodgkin lymphoma	MSS
SW620	Colorectal carcinoma	MSS
T47D	Breast carcinoma	MSS
U251/SNB-19	Glioma	MSS
UACC-257	Melanoma	MSS
UACC-62	Melanoma	MSS
UO-31	Renal cell carcinoma	MSS

Table 3-2: Microsatellite instability (MSI) status of the NCI-60. Legend: MSI- microsatellite instability, MSS- microsatellite stable, MSI-H- high microsatellite instability, MSI-L- low microsatellite instability. Cell lines were classified as MSS if none of five markers showed microsatellite instability. Cell lines were classified as MSI-H if two or more of five markers showed instability and MSI-L if one of five markers showed signs of instability. MSI status for LOXIMVI melanoma cell line is not available (NA). Adapted from the Cancer Genome Project website (http://www.sanger.ac.uk/genetics/CGP/MSI/msi_page.shtml).

3.3.2 *CTNNB1*

CTNNB1 is a dominantly acting cancer gene mutated mainly by missense mutations (COSMIC). In the NCI-60, one variant of *CTNNB1* was identified in a colorectal cancer cell line, HCT-116 (Table 3-3). The mutation of *CTNNB1* is characterized by a heterozygous in-frame deletion of three nucleotides corresponding to a serine residue at codon 45 of the reading frame. Codon 45 is most frequently affected by missense amino acid substitutions in cancer (COSMIC). There is no evidence of *APC* sequence variants in HCT-116. Therefore the presence of the *CTNNB1* in-frame deletion probably confers the same activation of the WNT pathway as a truncating *APC* variant and is consistent with what has previously been reported.

Cell line	Tumor type	*CTNNB1* variant identified	Type
HCT-116	Colorectal carcinoma	Het c.133_135 delTCT p.S45 del	LOM

Table 3-3. Mutation of CTNNB1 identified in the NCI-60 cell lines.

3.3.3 *MADH4*

Three variants of *MADH4* were identified in each of three NCI-60 cell lines (Table 3-4). *MADH4* is a tumor suppressor gene and the variants of *MADH4* are characterized by homozygous deletions, nonsense and frame-shifts resulting in a truncated protein product. The LOMs occurred in two colorectal cell lines with homozygous truncating variants. A heterozygous truncating variant was identified in one ovarian cell line.

Cell line	Tumor type	*MADH4* variants identified	Type
COLO-205	Colorectal carcinoma	Hom del exon1-6	LOM
HT-29	Colorectal carcinoma	Hom c.931C>T p.Q311X	LOM
IGROV-1	Ovarian carcinoma	Het c.692delG p.G231fsX10	TOV

Table 3-4. Mutations/variants of MADH4 identified in the NCI-60 cell lines.

***MADH4* cDNA sequence:**

A. *MADH4* IGROV-1 deletion of nucleotide 692:

gccagtatactgggggg**g**cagccatagtgaaggactgttgcaga**ta**
A S I L G G S H S E G L L Q I
gcatcagggcctcagccaggacagcagcagaatggatttactggt
A S G P Q P G Q Q Q N G F T G

Figure 3-3. Illustration of a nucleotide sequence change in MADH4 of the IGROV-1 ovarian cancer cell line. A) IGROV-1 a microsatellite unstable (MSI+) ovarian cancer cell line harbors a deletion of a guanine at nucleotide position 692 (bold red). This deletion occurs in a stretch of six guanine nucleotides (grey highlight) and causes a frame-shift resulting in a truncated protein product (black bold).

Mutations of *MADH4* occur in approximately 10% of colorectal cancers and are characterized by missense amino acid substitutions, frame-shifts, and homozygous deletions (COSMIC). Two of seven (29%) colorectal cancer cell lines of the NCI-60 have mutations of *MADH4*, slightly higher perhaps than that reported in primary tumors. *MADH4* is mutated in 7% of ovarian cancers and the mutations are mostly characterized by missense amino acid

substitutions and truncations (COSMIC). Therefore the presence of a frame-shift mutation of *MADH4* in IGROV-1 may conceivably be oncogenic. On the other hand, IGROV-1 is microsatellite unstable and the frame-shift mutation occurs in a polyguanine repeat sequence (Figure 3-3). Therefore the heterozygous mutation of *MADH4* in IGROV-1 may be a reflection the microsatellite instability and may not be cancer causing.

3.3.4 HRAS

One variant of *HRAS* was identified in a breast carcinoma cell line of the NCI-60, Hs-578T (Table 3-5). *HRAS* is a dominantly acting cancer gene and the variant of *HRAS* is characterized by a missense amino acid substitution affecting codon 12 of the reading frame. However, mutations of *HRAS* are generally rare in breast cancers (COSMIC). Sequencing of *HRAS* in 240 breast cancer tumors and cell lines reveal that ~1% of breast cancers harbor *HRAS* mutations (COSMIC).

Cell line	Tumor type	*HRAS* variants identified	Type
Hs-578-T	Breast carcinoma	Het c.35G>A p.G12D	LOM

Table 3-5. Mutation of HRAS identified in the NCI-60 cell lines.

The finding of an *HRAS* mutation in the NCI-60 breast cancers is therefore slightly surprising and potentially contrary to what has been seen in primary tumors. This may call into question the identity of Hs-578T as a breast cancer cell line. Gene expression analysis of the NCI-60 cell lines revealed that the panel of breast cancer cell lines are heterogeneous (Scherf

et al. 2000). Of the seven breast cancer cell lines in the NCI-60 two, MCF7 and T47D, cluster together (Figure 3-4). However, the remainder of the breast cancer cell lines are less similar to one another and do not cluster at all. In fact, Hs-578T clusters tightly with SF-559, a glioma cell line (Figure 3-4). Previously published data on mutations in *HRAS* (Koo et al. 1996) for the NCI-60 cell lines are consistent with those in this study.

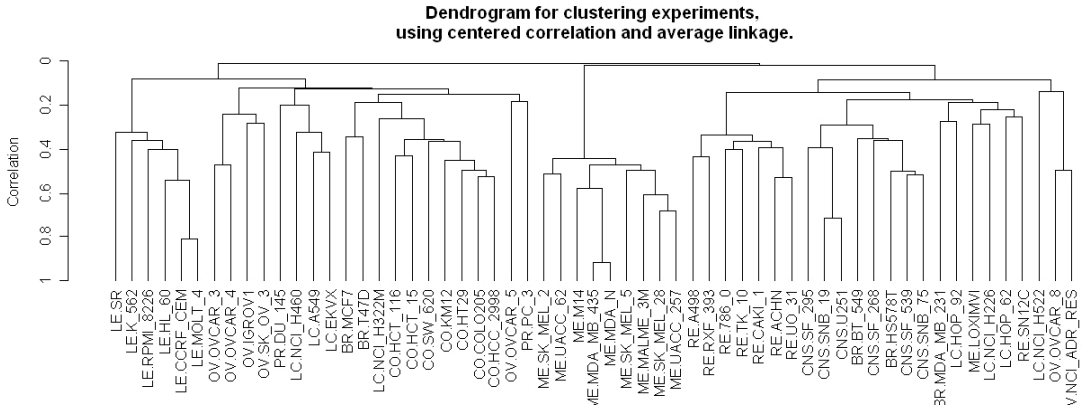

Figure 3-4: A clustered dendrogram of the NCI-60 cell lines using Affymetrix U133A gene expression data.

3.3.5 KRAS

Twelve variants of *KRAS* were identified in each of twelve cell lines of the NCI-60 (Table 3-6). *KRAS* is a dominantly acting cancer gene and the mutations of *KRAS* are characterized by missense amino acid substitutions affecting codons 12, 13, 61 and 146. Of the eleven cell lines with *KRAS* mutations, four are lung carcinoma cell lines, four are colon carcinoma cell lines and the remaining four *KRAS* mutations were identified in a leukemia, breast carcinoma, ovarian carcinoma and myeloma.

Mutations of *KRAS* occur in approximately 18% of lung cancers (COSMIC). I report that four of eight (50%) of the lung cancer cell lines in the NCI-60 panel harbor *KRAS* mutations. Three of the *KRAS* mutations in the lung cancer cell lines affect codon 12 and one *KRAS* mutation affects codon 61. Less than 1% of lung cancers with a *KRAS* mutation harbor amino acid substitutions of codon 61 (COSMIC). The presence of a codon 61 KRAS mutation in one of the eight lung cancer samples of the NCI-60 is therefore unusual.

KRAS mutations occur in 32% of colorectal cancers (COSMIC). I observe that four of seven (57%) colorectal cancer cell lines of the NCI-60 harbor *KRAS* mutations. Most *KRAS* mutations in colorectal cancer occur at codons 12 and 13. *KRAS* mutations in colorectal cancer also occur at codons 19, 61 and were recently identified as a recurrent event at codon 146 (Edkins et al. 2006, COSMIC). One of the seven colorectal cell lines of the NCI-60 harbors an amino acid substitution at *KRAS* codon 146.

KRAS mutations occur in 12% of acute lymphoblastic leukemias (ALL) (COSMIC). *KRAS* mutations in ALL generally occur at codon 12 and I identified one of two lymphoblastic leukemia cell lines of the NCI-60 with a codon 12 *KRAS* mutation.

KRAS mutations occur in 5% of breast carcinomas (COSMIC). Most of the *KRAS* mutations in breast carcinoma occur at codon 12. Approximately

1% of *KRAS* mutations in breast carcinomas occur at codon 13 with an amino acid change from glycine to aspartic acid. I identified one of five breast carcinoma cell lines of the NCI-60 with a G13D glycine to aspartic acid *KRAS* mutation. Therefore, again, the pattern of *RAS* gene mutation in breast cancers is unusual, and may call into question the underlying tissue of origin of some of the cell lines.

KRAS mutations occur in 13% of ovarian carcinomas (COSMIC). The mutations of *KRAS* in ovarian carcinoma usually affect codon 12 and I identified one of seven ovarian carcinoma cell lines of the NCI-60 with a G12V *KRAS* mutation.

KRAS mutation occurs in 17% of myelomas (COSMIC). The mutations of *KRAS* in myeloma usually affect codon 12 and I identified a KRAS codon 12 mutation in the single myeloma cell line of the NCI-60 panel.

Previously published data on mutations in *KRAS* (Koo et al. 1996) for the NCI-60 cell lines are consistent with those in this study.

Cell line	Tumor type	*KRAS* variants identified	Type
A549	Lung carcinoma	Hom c.34G>A p.G12S	LOM
CCRF-CEM	Acute lymphoblastic leukemia	Het c.35G>A p.G12D	LOM
HCC2998	Colon carcinoma	Het c.436G>A p.A146T	LOM
HCT-116	Colon carcinoma	Het c.38G>A p.G13D	LOM
HCT-15	Colon carcinoma	Het c.38G>A p.G13D	LOM
HOP-62	Lung adenocarcinoma	Het c.34G>T p.G12C	LOM
MDA-MB-231	Breast carcinoma	Het c.38G>A p.G13D	LOM
NCI-H23	Lung adenocarcinoma	Het c.34G>T p.G12C	LOM
NCI-H460	Lung large cell carcinoma	Hom c.183A>T p.Q61H	LOM
OVCAR-5	Ovarian carcinoma	Hom c.35G>T p.G12V	LOM
RPMI-8226	Myeloma	Het c.35G>C p.G12A	LOM
SW620	Colon carcinoma	Hom c.35G>T p.G12V	LOM

Table 3-6. Mutations of KRAS identified in the NCI-60 cell lines.

3.3.6 NRAS

Three mutations of *NRAS* were identified in each of three NCI-60 cell lines (Table 3-7). *NRAS* is a dominantly acting cancer gene and the mutations of *NRAS* are characterized by missense amino acid substitutions affecting codons 12, 13 and 61.

NRAS mutations occur in 22% of melanomas (COSMIC). The mutations of *NRAS* in melanoma most commonly occur at codon 61. I identified an *NRAS* mutation at codon 61 in one of nine (11%) melanoma cell lines of the NCI-60. The melanoma cell line with the codon 61 *NRAS* mutation does not have a *BRAF* mutation.

NRAS mutations occur in 38% of acute lymphoblastic leukemia (ALL) (COSMIC). The mutations of *NRAS* in ALL occur most frequently at codon 12. I identified an NRAS codon 12 mutation in one of two (50%) ALL cell lines of the NCI-60.

NRAS mutations occur in 16% of acute myeloid leukemia (AML) (COSMIC). The mutations of *NRAS* in AML occur most frequently at codon 12. I identified an NRAS codon 12 mutations in the only AML cell line of the NCI-60.

Previously published data on mutations in *NRAS* (Koo et al. 1996) for the NCI-60 cell lines are consistent with those in this study.

Cell line	Tumor type	*NRAS* variants identified	Type
HL-60	Acute myeloid leukemia	Het c.182A>T p.Q61L	LOM
MOLT-4	Acute lymphoblastic leukemia	Het c.34G>T p.G12C	LOM
SKMEL-2	Melanoma	Hom c.182A>G p.Q61R	LOM

Table 3-7. Mutations of NRAS identified in the NCI-60 cell lines.

3.3.7 BRAF

Eleven mutations of *BRAF* were identified in each of eleven NCI-60 cell lines (Table 3-8). *BRAF* is a dominantly acting cancer gene and mutations are characterized by missense amino acid substitutions mainly at codon 600. Ten of the eleven mutations I detected occur at codon 600. Eight of the eleven cell lines harboring *BRAF* mutations are melanomas, two are colorectal cancer cell lines and one is a breast cancer.

BRAF mutations occur in 47% of melanomas (COSMIC). The mutations of *BRAF* in melanoma occur most frequently at codon 600. I

identified BRAF codon 600 mutations in eight of nine (89%) melanoma cell lines of the NCI-60.

BRAF mutations occur in 15% of colorectal cancers (COSMIC). The mutations of *BRAF* in colorectal cancer occur most frequently at codon 600. I identified codon 600 mutations in two of seven (29%) colorectal cancer cell lines of the NCI-60. It has been reported that BRAF mutations tend to occur in microsatellite unstable colorectal cancers (Rajagopalan et al. 2002). However, the two colorectal cancers of the NCI-60 with BRAF mutation are not microsatellite unstable.

All of the cell lines with BRAF mutation at codon 600 do not harbor coincident *RAS* mutations. This mirrors the pattern of *BRAF* mutation in primary tumors where V600E *BRAF* mutation almost never co-occurs with a *RAS* mutation (Davies et al. 2002, Thomas et al. 2007).

BRAF mutations are reported to occur in 3% of breast carcinomas (COSMIC). I identified a codon 464 mutation in one of five (20%) breast cancer cell lines of the NCI-60. Codon 464 *BRAF* mutations within the P loop have previously been found in human cancer, but are relatively rare (COSMIC). It is however, unusual to find a *BRAF* mutation in breast cancer. MDA-MB-231, the breast cancer cell line of the NCI-60 with a codon 464 BRAF mutation also harbors a coincident *KRAS* mutation. It may be that MDA-MB-231 is a rare type of breast cancer. It may also be that MDA-MB-231 is not a breast cancer cell line. I find that MDA-MB-231 clusters tightly

with a lung large cell carcinoma line, HOP-92 (Figure 3-4). Interestingly, it is more common to find non-codon 600 BRAF mutations and coincident RAS mutations in lung cancer than in melanoma or colorectal cancers (Davies et al. 2002).

Finally, by re-sequencing of cancer genes and genotyping of 10,000 SNPs in the NCI-60 panel (http://www.sanger.ac.uk/genetics/CGP/Genotyping/nci60.shtml), I find that melanoma cell line, M14 and 'breast' cancer cell line, MDA-MB-435 share 97% genotype similarity and identical mutations in *BRAF*, *CDKN2A*, and *TP53*. MDA-MB-435 has previously been labeled as a breast cancer cell line. Expression analysis revealed that MDA-MB-435 clustered tightly with another melanoma cell line, MDA-N, of the NCI-60 (Ross et al. 2000). Due to the previously reported expression analysis, its overwhelming genotypic similarity to M14 another melanoma cell line of the NCI-60, and the combination of mutations in *BRAF* and *CDKN2A*, MDA-MB-435 is almost certainly a melanoma cell line.

Cell line	Tumor type	*BRAF* variants identified	Type
COLO-205	Colorectal carcinoma	Het c.1799T>A p.V600E	LOM
HT-29	Colorectal carcinoma	Het c.1799T>A p.V600E	LOM
LOXIMVI	Melanoma	Het c.1799T>A p.V600E	LOM
M14	Melanoma	Het c.1799T>A p.V600E	LOM
MDA-MB-435	Melanoma	Het c.1799T>A p.V600E	LOM
MALME-3M	Melanoma	Het c.1799T>A p.V600E	LOM
MDA-MB-231	Breast carcinoma	Het c.1391G>T p.G464V	LOM
SKMEL-28	Melanoma	Hom c.1799T>A p.V600E	LOM
SKMEL-5	Melanoma	Het c.1799T>A p.V600E	LOM
UACC-257	Melanoma	Het c.1799T>A p.V600E	LOM
UACC-62	Melanoma	Hom c.1799T>A p.V600E	LOM

Table 3-8. Mutations of BRAF identified in the NCI-60 cell lines.

3.3.8 PIK3CA

Seven mutations of *PIK3CA* were identified in each of seven cell lines of the NCI-60 (Table 3-9). Of the seven cell lines, three are colon, two are breast lines, one is lung and one is an ovarian cancer cell line. *PIK3CA* is a dominantly acting cancer gene and the mutations of *PIK3CA* are characterized by heterozygous missense amino acid substitutions affecting most frequently codons 449, 545 and 1047. Previously published data on mutations of *PIK3CA* (Whyte and Holbeck et al. 2006) for the NCI-60 cell lines are consistent with those in this study.

PIK3CA mutation occurs in 23% of colorectal cancers (COSMIC). I identified *PIK3CA* mutations at codon 449, 545 and 1047 in three of seven (43%) colorectal cell lines of the NCI-60.

PIK3CA mutation occurs in 27% of breast carcinomas (COSMIC). I identified *PIK3CA* mutations at codon 545 and 1047 in two of five (40%) breast carcinoma cell lines of the NCI-60.

PIK3CA mutation occurs in approximately 5% of lung large cell carcinomas (COSMIC). I identify a codon 545 *PIK3CA* mutation in one of two (50%) lung large cell carcinomas of the NCI-60.

PIK3CA mutation occurs in 9% of ovarian carcinomas (COSMIC). I identified a codon 1047 PIK3CA mutation in one of seven (14%) ovarian carcinoma cell lines of the NCI-60.

Cell line	Tumor type	*PIK3CA* variants identified	Type
HCT-116	Colon carcinoma	Het c.3140A>G p.H1047R	LOM
HCT-15	Colon carcinoma	Het c.1633G>A p.E545K	LOM
HT-29	Colon carcinoma	Het c.1345C>A p.P449T	LOM
MCF7	Breast carcinoma	Het c.1633G>A p.E545K	LOM
NCI-H460	Lung large cell carcinoma	Het c.1633G>A p.E545K	LOM
SK-OV-3	Ovarian carcinoma	Het c.3140A>G p.H1047R	LOM
T47D	Breast carcinoma	Het c.3140A>G p.H1047R	LOM

Table 3-9. Mutations of PIK3CA identified in the NCI-60 cell lines.

3.3.9 PTEN

Thirteen mutations of *PTEN* were identified in twelve cell lines of the NCI-60 (Table 3-10). *PTEN* is a tumor suppressor gene and the mutations of *PTEN* are characterized by homozygous deletions, homozygous nonsense and frame-shift mutations resulting in a truncated protein product. The mutations occurred in four glioma cell lines, two lymphoblastic leukemias, two renal cell carcinomas, and one each of an ovarian carcinoma, colorectal carcinoma, prostate cancer, and melanoma.

PTEN mutations occur in 21% of gliomas (COSMIC). I identified homozygous nonsense and frame-shift mutations of *PTEN* in four of six (67%) glioma cell lines of the NCI-60. The frequency of *PTEN* mutation in glioma cell lines of the NCI-60 therefore appears slightly higher than that observed in primary tumors.

Cell line	Tumor type	*PTEN* variants identified	Type
786-0	Renal cell carcinoma	Hom c.445C>T p.Q149X	LOM
CCRF-CEM	Acute lymphoblastic leukemia	Hom c. del 80-492 p.?	LOM
IGROV-1	Ovarian carcinoma	Het c.955_958delACTT p.T319fsX1	TOV
KM12	Colorectal carcinoma	Het c.385G>T p.G129X, Het c.800del A p.K267fsX9	LOM
MOLT-4	Acute lymphoblastic leukemia	Hom c.800delA p.K267fsX9	LOM
PC-3	Prostate carcinoma	Hom c.165-1026 del 862 p.?	LOM
RXF393	Renal cell carcinoma	Hom c.1_164 del 164 p.?	LOM
SF-295	Glioma	Hom c.697C>T p.R233X	LOM
SF539	Glioma	Hom c.1-1026 del 1026 p.?	LOM
SNB-19	Glioma	Hom c.723_724insTT p.E242fsX15	LOM
U251	Glioma	Hom c.723_724insTT p.E242fsX15	LOM
UACC-62	Melanoma	Hom c.741_742ins A p.P248fsX5	LOM

Table 3-10. Mutations/variants of PTEN identified in the NCI-60 cell lines.

PTEN mutations occur in 6% of hematopoeitic and lymphoid cancers (COSMIC). I identified a homozygous deletion and a frame-shift mutation of *PTEN* in the two acute lymphoblastic leukemia cell lines of the NCI-60.

PTEN mutations occur in 13% of renal cell carcinomas (COSMIC). I identified a nonsense mutation and homozygous deletion of *PTEN* in two of eight (25%) renal cell carcinoma cell lines of the NCI-60.

PTEN mutations occur in 14% of prostate carcinomas (COSMIC). I identified a homozygous deletion of *PTEN* in one of two (50%) prostate carcinoma cell lines of the NCI-60. The frequency of *PTEN* mutation in the prostate carcinoma cell lines of the NCI-60 is higher than that seen in primary tumors.

PTEN mutations occur in 18% of melanomas (COSMIC). I identified a frame-shift mutation of *PTEN* in one of nine (11%) melanoma cell lines of the NCI-60. The frequency of *PTEN* mutations in melanoma cell lines of the NCI-60 is therefore consistent with that observed in primary tumors.

***PTEN* cDNA sequence**:

A. *PTEN* KM12 sequence change of nucleotide 385 from G>T:

gcaattcactgtaaagctggaaag**g**g**a**cgaactggtgtaatgata
A I H C K A G K G R T G V M I

B. *PTEN* KM12 deletion of nucleotide 800:

gagttcttccacaaacagaacaagatgctaaaaaaggacaaaatg
E F F H K Q N K M L K K D K M
tttcacttttggg**taa**atacattcttcataccaggaccagaggaa
F H F W V N T F F I P G P E E

C. *PTEN* IGROV-1 deletion of nucleotides 9555-9558:

ctagtacttactt**taa**caaaaaatgatcttgacaaagcaaataaa
L V L T L T K N D L D K A N K

Figure 3-5: Illustration of a sequence changes in PTEN of KM12 colorectal cell line and IGROV-1 ovarian cancer cell line. A) KM12 a microsatellite unstable (MSI+) colorectal cell line harbors a heterozygous nucleotide change at position 385 from a G>T (bold red) resulting in a stop (tga) codon (bold black). B) KM12 harbors a heterozygous deletion of adenine at nucleotide position 800 (bold red) in a stretch of six adenine nucleotides (grey highlight) resulting in a frame-shift and a premature stop (taa) codon (bold black). C) IGROV-1 a microsatellite unstable (MSI+) ovarian cancer cell line harbors a deletion of four nucleotides at positions 9555 to 9558 (bold red) in a tetra-

nucleotide repeat sequence (grey highlight) resulting in a frame-shift and a premature stop (taa) codon (bold black).

Three truncating variants of *PTEN* were identified in two microsatellite unstable cell lines, IGROV-1 and KM12. IGROV-1 harbors a heterozygous frame-shift truncating variant of *PTEN*. The frame-shift mutation occurred in a repeat sequence consistent with the microsatellite instability of the cell line (Figure 3-5). Therefore, the truncating variant of *PTEN* is most likely not contributing to tumorigenesis in IGROV-1.

The other microsatellite unstable cell line, KM12, harbors two heterozygous truncating variants of *PTEN*. One of the variants involves a G>T nucleotide change resulting in the coding of a stop codon. The G>T change is not typical of the C>T transition predominant in microsatellite unstable cell lines (Greenman et al. 2007). Therefore, it is perhaps slightly more likely that the G>T transversion resulting in a stop codon of *PTEN* in KM12 is contributing to tumorigenesis. The second variant of *PTEN* in KM12 involves a deletion of an adenine nucleotide in a stretch of six adenine nucleotides. This frame-shift mutation results in a premature truncation codon and is consistent with the sort of truncating variants seen in MSI+ cancer cell lines. Despite the fact that KM12 is MSI+, the presence of two heterozygous truncating variants of *PTEN* suggests that perhaps both alleles of *PTEN* are inactivated and contributing to tumorigenesis. However, it is not clear from these data whether the *PTEN* variants affect one or both alleles.

By re-sequencing of cancer genes and genotyping of 10,000 SNPs in the NCI-60 panel (http://www.sanger.ac.uk/genetics/CGP/Genotyping/nci60.shtml), I find that the glioma cell lines, SNB-19 and U251 share 99% genotype similarity and identical *PTEN* and *TP53* mutations. Therefore SNB-19 and U251 are derived from the same individual and are likely the same cell line.

3.3.10 STK11

Five mutations of *STK11* were identified in each of five cell lines of the NCI-60 (Table 3-11). *STK11* is a tumor suppressor gene and the mutations of *STK11* were identified in three lung carcinomas, one prostate and one lymphoblastic leukemia cell line. The mutations of *STK11* are characterized by nonsense and frame-shift mutations resulting in a truncated protein product. *STK11* mutations occur in 10% of lung cancers (COSMIC). I identified nonsense mutations of *STK11* in three of eight (38%) lung cancer cell lines of the NCI-60.

STK11 mutations have never been reported to occur in prostate cancers. However, I have identified a homozygous frame-shift mutation of *STK11* in a prostate cancer cell line of the NCI-60. DU-145 has been reported to be null for STK11 protein expression (Yun et al. 2005). It will be of interest to extend that observation to a set of primary prostate cancers to determine the prevalence of *STK11* inactivation in this common tumor type. It may be that *STK11* inactivation is necessary for tumorigenesis in a subset of prostate carcinomas. However, DU145 is a microsatellite unstable cell line

with a frameshift mutation (Figure 3-6). Therefore, it is possible that the variant in STK11 is a passenger mutation and not likely oncogenic. On the other hand, given that STK11 protein expression in the cell line is absent, probably due to the truncating variant I have identified, it is equally possible that the variant is contributing to oncogenesis.

STK11 mutations in acute lymphoblastic leukemia have also never been reported. I have identified a heterozygous nonsense mutation in one lymphoblastic leukemia cell line. However, MOLT-4, the lymphoblastic leukemia cell line with a nonsense *STK11* mutation is also microsatellite unstable. It has been reported that C to T transitions are a common phenomenon in microsatellite unstable cancers (Greenman et al. 2007). The nonsense mutation of *STK11* in MOLT-4 involves a C to T transition at nucleotide 640. Therefore, it is possible that the nonsense mutation of *STK11* in MOLT-4 is due to the microsatellite instability and, hence, the elevated base substitution mutation rate of the cell line. Therefore, the nonsense mutation of *STK11* in MOLT-4 is classified as tentatively oncogenic.

Cell line	Tumor type	*STK11* variants identified	Type
A549	Lung carcinoma	Hom c.109C>T p.Q37X	LOM
DU145	Prostate carcinoma	Hom c.532_536delAAGCC p.K178fsX86	LOM
MOLT-4	Acute lymphoblastic leukemia	Het c.640C>T p.Q214X	TOV
NCI-H23	Lung adenocarcinoma	Hom c.996G>A p.W332X	LOM
NCI-H460	Lung large cell carcinoma	Hom c.109C>T p.Q37X	LOM

Table 3-11. Mutations/variants of STK11 identified in the NCI-60 cell lines.

***STK11* cDNA sequence:**

A. *STK11* DU145 sequence change deletion nucleotides 532_536:

tacctgcatagccagggcattgtgcacaaggacatc**aagcc**gggg
Y L H S Q G I V H K D I K P G

B. *STK11* MOLT-4 sequence change nucleotide 640 C>T:

cggaccagc**c****ag**ggctccccggctttccagccgcccgagattgcc
R T S Q G S P A F Q P P E I A

Figure 3-6: Illustration of a nucleotide sequence change in STK11 of the MOLT-4 acute lymphoblastic leukemia cancer cell line. A) DU145 a microsatellite unstable (MSI+) prostate carcinoma cell line harbors a deletion of nucleotides 532 to 536 (AAGCC) (bold black) resulting in a frame-shift and a truncated protein product. B) MOLT-4 a microsatellite unstable (MSI+) acute lymphoblastic leukemia cancer cell line harbors a C>T transition at nucleotide position 640 (bold red) resulting in a stop codon (bold black).

3.3.11 VHL

Two mutations of *VHL* were identified in each of two renal cell carcinoma cell lines of the NCI-60 (Table 3-12). *VHL* is a tumor suppressor gene and the mutations of *VHL* are characterized by frame-shift mutations resulting in a truncated protein product. *VHL* mutations occur in 43% of renal cancers (COSMIC). I identified homozygous frame-shift mutations in two of eight (25%) renal cell carcinoma cell lines of the NCI-60. The frequency of *VHL* mutations in renal cancer cell lines of the NCI-60 is similar or slightly lower than that observed in primary tumors.

Cell line	Tumor type	*VHL* variants identified	Type
786-0	Renal cell carcinoma	Hom c.311delG p.G105fsX55	LOM
A498	Renal cell carcinoma	Hom c.426_429delTGAC p.G144fsX14	LOM

Table 3-12. Mutations of VHL identified in the NCI-60 cell lines.

3.3.12 RB1

Four mutations of *RB1* were identified in each of four cell lines of the NCI-60 (Table 3-13). *RB1* is a tumor suppressor gene and the mutations of *RB1* are characterized by homozygous deletions, nonsense, and frame-shift mutations resulting in a truncated protein product. *RB1* mutations occur in 11% of breast cancers, 11% of prostate cancers, and approximately 10% of cancers of the nervous system (COSMIC). I identified a homozygous deletion of *RB1* in one of five (20%) breast carcinoma cell lines of the NCI-60, a homozygous nonsense *RB1* mutation in one of two (50%) prostate cancer cell lines, and a homozygous frame-shift mutation in one of six (17%) glioma cell lines.

However, mutations of *RB1* occur in less than 1% of colorectal cancers (COSMIC). I identified a heterozygous nonsense mutation of *RB1* in one colorectal cancer cell line (HCC2998) of the NCI-60. HCC2998 is a microsatellite stable cell line (MSI-). Therefore, the heterozygous truncating variant of *RB1* is not due to defective mismatch repair. The heterozygous truncating variant of *RB1* only inactivates one allele and is therefore classified as only tentatively oncogenic.

Cell line	Tumor type	*RB1* variants identified	Type
BT-549	Breast carcinoma	Hom c.265_607 del 343 p.?	LOM
DU145	Prostate carcinoma	Hom c.2143A>T p.K715X	LOM
HCC2998	Colon carcinoma	Het c.409G>T p.E137X	TOV
SF539	Glioma	Hom c.346_349delACTT p.T116fsX8	LOM

Table 3-13. Mutations/variants of RB1 identified in the NCI-60 cell lines.

3.3.13 CDKN2A

Thirty-eight mutations of *CDKN2A* were identified in 33 cell lines of the NCI-60. *CDKN2A* is a tumor suppressor gene, the mutations of *CDKN2A* occur in every tissue type represented in the NCI-60 cell lines, and are characterized by homozygous deletions, missense, nonsense, splice site and frame-shift mutations resulting in a truncated protein product. *CDKN2A* mutations occur in 15% of human cancers (COSMIC) including 24% of melanomas, 22% of gliomas, 15% of lung cancers, 12% of renal cancers and 5% of breast cancers (COSMIC). I identified mutations of *CDKN2A* in 33 of 59 (56%) of the NCI-60 human cancer cell lines including 67% of melanomas, 66% of gliomas, 56% of lung cancer cell lines, 75% of renal cancer cell lines, and 60% of breast cancer cell lines in the NCI-60. Overall, the frequency of *CDKN2A* mutations in the NCI-60 is higher than that observed in primary tumors.

It has previously been reported that *CDKN2A* deletions are more frequently observed in cancer cell lines compared to primary tumors (Kamb et al. 1994). This finding may be explained by a possible selection bias for primary tumors with *CDKN2A* mutations. Perhaps primary tumors with

CDKN2A mutations are more likely to take to growing in tissue culture than tumors without *CDKN2A* mutations.

CDKN2A gene gives rise to two distinct transcripts with overlapping but different open reading frames known as *p16* and *p14*. All of the homozygous deletion mutations of *CDKN2A* I identified in the NCI-60 cell lines encompass both transcripts. I identified a small insertion that causes a translational frame-shift variant of *p16* only, in the colorectal cancer cell line HCT-116. There was no predicted mutation of p14 because the variant occurred outside the coding region of p14. I also identified a base substitution that caused a missense amino acid substitution variant of both *p16* and *p14* in the prostate cancer cell line, DU145. In the AML cell line, HL60, there is a base substitution variant that causes a nonsense codon in p16 but a missense amino acid substitution in p14. Overall, our mutational data do not clearly indicate whether p16, p14 or both are primary targets of mutation in these genes, although they marginally favour p16. However, other data indicate that base substitution variants found in *CDKN2A* generate nonsense codons much more frequently in p16 than in p14, indicating that p16 is the major target.

Cell line	Tumor type	*CDKN2A* variants identified	Type
786-0	Renal cell carcinoma	Hom c.1_150 del150 p.?	LOM
A498	Renal cell carcinoma	Hom c.1_471 del471 p.?	LOM
A549	Lung carcinoma	Hom c.1_471 del471 p.?	LOM
ACHN	Renal cell carcinoma	Hom c.1_471 del471 p.?	LOM
CAKI-1	Renal cell carcinoma	Hom c.1_471 del471 p.?	LOM
CCRF-CEM	Acute lymphoblastic leukemia	Hom c.1_471 del471 p.?	LOM

DU145	Prostate carcinoma	Hom c.250G>T p.D84Y (p16), Hom c. 416G>T p. R139L (p14)	LOM
HCT-116	Colorectal carcinoma	Het c.68_69insG p.R24fsX20 (p16)	LOM
HL60	Acute myeloid leukemia	Hom c.238C>T p.R80X (p16) Hom c.338C>T p.P113L (p14)	LOM
HOP-62	Lung adenocarcinoma	Hom c.1_471 del471 p.?	LOM
HOP-92	Lung large cell carcinoma	Hom c.1_471 del471 p.?	LOM
Hs-578-T	Breast carcinoma	Hom c.1_471 del471 p.?	LOM
K-562	Chronic myeloid leukemia	Hom c.1_471 del471 p.?	LOM
LOXIMVI	Melanoma	Hom c.1_471 del471 p.?	LOM
M14	Melanoma	Het c.150+2 T>C p.?, Het c.456-+24 AGgtgaggactgatgatctgagaattt>C p.?	LOM
MALME-3M	Melanoma	Hom c.1_471 del471 p.?	LOM
MCF7	Breast carcinoma	Hom c.1_471 del471 p.?	LOM
MDA-MB-231	Breast carcinoma	Hom c.1_471 del471 p.?	LOM
MDA-MB-435	Melanoma	Het c.150+2 T>C p.?, Het c.456-+24 AGgtgaggactgatgatctgagaattt>C p.?	LOM
MOLT-4	Acute lymphoblastic leukemia	Hom c.1_471 del471 p.?	LOM
NCI-H226	Lung squamous cell carcinoma	Hom c.1_150 del150 p.?	LOM
NCI-H460	Lung large cell carcinoma	Hom c.1_457 del457 p.?	LOM
OVCAR-5	Ovarian carcinoma	Hom c.1_471 del471 p.?	LOM
RXF393	Renal cell carcinoma	Hom c.1_471 del471 p.?	LOM
SF-268	Glioma	Hom c.1_471 del471 p.?	LOM
SF-295	Glioma	Hom c.1_471 del471 p.?	LOM
SK-MEL-5	Melanoma	Hom c.1_471 del471 p.?	LOM
SK-OV-3	Ovarian carcinoma	Hom c.1_457 del457 p.?	LOM
SNB-19	Glioma	Hom c.1_471 del471 p.?	LOM
SR	Non Hodgkin lymphoma	Hom c.1_471 del471 p.?	LOM
U251	Glioma	Hom c.1_471 del471 p.?	LOM
UACC-62	Melanoma	Hom c.1_471 del471 p.?	LOM
UO-31	Renal cell carcinoma	Hom c.1_471 del471 p.?	LOM

Table 3-14. Mutations of CDKN2A identified in the NCI-60 cell lines.

3.3.14 TP53

Forty-seven mutations of *TP53* were identified in 43 cell lines of the NCI-60 (Table 3-15). *TP53* is a tumor suppressor gene and mutations of *TP53* occurred in every tissue type represented in the NCI-60 cell lines. *TP53*

mutations are characterized by missense, nonsense, frame-shift, and homozygous deletions resulting in altered forms of the protein product. The *TP53* mutations occur most frequently in the sequence specific binding domain and oligomerization domain (Figure 3-8). *TP53*, the most commonly mutated gene in cancer, was altered in 64% (38/59) of the NCI-60 cell lines. *TP53* mutations occur frequently in ovarian cancers (48%), colorectal cancers (44%), lung cancers (38%), skin cancers (35%), bladder cancers (28%), gliomas (27%), and breast cancers (25%) (Petitjean et al. 2007; IARC). TP53 mutations occur less frequently in prostate cancers (17%) and hematopoietic cancers (13%). (Petitjean et al. 2007, IARC).

The frequency of *TP53* mutations in the NCI-60 cell lines adequately reflects that seen in the most common primary tumors with *TP53* mutations. However, with respect to the frequency of *TP53* mutations in cancers in which *TP53* mutation is rare, the NCI-60 cell line panel is not representative. For example, both prostate (100%) cancer cell lines of the NCI-60 harbor *TP53* mutations and 57% (4/7) of the hematopoietic cell lines of the NCI-60 harbor TP53 mutations.

Previously published data on mutations in *TP53* for the NCI-60 cell lines are not completely consistent with those in this study. With respect to the previously published *TP53* sequence analysis by O'Connor (O'Connor et al. 1997), different results were obtained for 9 of the 59 cell lines. Some are annotation differences in the *TP53* data: HS578T has a p.V157F mutation here but p.D157E reported, RPMI-8226 is p.E285K here but has a previous

annotation of p.E285L, and SK-MEL-28 is p.L145R here rather than p.C145V (O'Connor et al. 1997). In addition, in our analysis, MOLT-4 has a heterozygous *TP53* nonsense mutation (p.R306X) in genomic DNA but no detectable mutation at the cDNA level in the previous study. In this instance it is plausible that the mutant *TP53* transcript in MOLT-4 undergoes nonsense-mediated decay and therefore is not detectable in cDNA.

Cell line	Tumor type	*TP53* variants identified	Type
786-0	Renal cell carcinoma	Het c.832C>G p.P278A, Het c.A560-2A>G p.?	LOM
BT-549	Breast carcinoma	Hom c.747G>C p.R249S	LOM
CCRF-CEM	Acute lymphoblastic leukemia	Het c.743G>A p.R248Q, Het c.524G>A p.R175H	LOM
COLO-205	Colorectal carcinoma	Hom c.308_333>TA p.Y103fsX37	LOM
DU-145	Prostate carcinoma	Het c.820G>T p.V274F	LOM
EKVX	Lung adenocarcinoma	Hom c.609_610GG>TT p.E204X	LOM
HCC2998	Colorectal carcinoma	Het c.637C>T p.R213X	TOV
HCT-15	Colorectal carcinoma	Het c.C1101-2A>C p.?, Het c.722C>T p.S241F	LOM
HL60	Acute myeloid leukemia	Hom deletion (*)	LOM
HOP-62	Lung adenocarcinoma	Hom c.G673-2A>G p.?	LOM
HOP-92	Lung large cell carcinoma	Hom c.524G>T p.R175L	LOM
Hs-578-T	Breast carcinoma	Hom c.469G>T p.V157F	LOM
HT-29	Colorectal carcinoma	Hom c.818G>A p.R273H	LOM
IGROV-1	Ovarian carcinoma	Het c.377A>G p.Y126C	LOM
K-562	Chronic myeloid leukemia	Hom c.406_407insC p.Q136fsX13	LOM
KM12	Colorectal carcinoma	Het c.215delC p.P72fsX51	TOV
M14	Melanoma	Het c.797G>A p.G266E	LOM
MDA-MB-231	Breast carcinoma	Hom c.839G>A p.R280K	LOM
MDA-MB-435	Melanoma	Het c.797G>A p.G266E	LOM
MOLT-4	Acute lymphoblastic leukemia	Het c.916C>T p.R306X	TOV
NCI/ADR-RES	Ovarian carcinoma	Hom c.376-1G>A p.?	LOM
NCI-H23	Lung adenocarcinoma	Hom c.738G>C p.M246I	LOM
NCI-H322M	Lung bronchoalveolar carcinoma	Hom c.743G>T p.R248L	LOM
NCI-H522	Lung adenocarcinoma	Hom c.572delC p.P191fsX56	LOM
OVCAR-3	Ovarian carcinoma	Hom c.743G>A p.R248Q	LOM
OVCAR-4	Ovarian carcinoma	Hom c.388C>G p.L130V	LOM
OVCAR-8	Ovarian carcinoma	Hom c.376-1G>A p.?	LOM
PC-3	Prostate carcinoma	Hom c.414delC	LOM

		p.K139fsX31	
RPMI-8226	Myeloma	Hom c.853G>A p.E285K	LOM
RXF393	Renal cell carcinoma	Hom c.524G>A p.R175H	LOM
SF-268	Glioma	Hom c.818G>A p.R273H	LOM
SF-295	Glioma	Hom c.743G>A p.R248Q	LOM
SF-539	Glioma	Hom c.1024delC p.R342fsX3	LOM
SK-MEL-2	Melanoma	Het c.733G>A p.G245S	LOM
SK-MEL-28	Melanoma	Hom c.435_436G>GT p.L145R	LOM
SK-OV-3	Ovarian carcinoma	Hom c.267delC p.S90fsX33	LOM
SN12C	Renal cell carcinoma	Hom c.1006G>T p.E336X	LOM
SNB-19	Glioma	Hom c.818G>A p.R273H	LOM
SNB-75	Glioma	Hom c.772G>A p.E258K	LOM
SW620	Colorectal carcinoma	Hom c.818G>A p.R273H, Hom c.925C>T p.P309S	LOM
T47D	Breast carcinoma	Hom c.580C>T p.L194F	LOM
TK10	Renal cell carcinoma	Het c.791T>G p.L264R	LOM
U251	Glioma	Hom c.818G>A p.R273H	LOM

*Table 3-15. Mutations/variants of TP53 identified in the NCI-60 cell lines. * homozygous deletion of TP53 in HL60 was previously reported.*

***TP53* cDNA sequence**:

A. *TP53* KM12 deletion of nucleotide 215:

gatgaagctcccagaatgccagaggctgctccccccgtggcccct
D E A P R M P E A A P P V A P
gcaccagcagctcctacaccggcggcccctgcaccagccccctcc
A P A A P T P A A P A P A P S
tggcccctgtcatcttctgtcccttcccagaaaacctaccagggc
W P L S S S V P S Q K T Y Q G
agctacggtttccgtctgggcttcttgcattctgggacagccaag
S Y G F R L G F L H S G T A K
tctg**tga**ct
S V T

B. *TP53* MOLT-4 nucleotide change 916 C>T:

ccagggagcactaag**cga**gcactgcccaacaacaccagctcctct
P G S T K R A L P N N T S S S

Figure 3-7: Illustration of nucleotide sequence changes in TP53 of the KM12 colorectal cancer cell line and the MOLT-4 acute lymphoblastic leukemia cancer cell line. A) KM12 a microsatellite unstable (MSI+) colorectal cancer cell line harbors a deletion of a cytosine at position 215 (bold red) within a sequence stretch of six cytosine nucleotides (grey highlight). This causes a frame-shift resulting in a premature stop (tga) codon (bold black). B) MOLT-4 a microsatellite unstable (MSI+) acute lymphoblastic leukemia cancer cell line harbors a C>T transition at nucleotide position 916 (bold red) resulting in a stop codon (bold black).

I identified three TOVs of *TP53* in three NCI-60 cell lines. HCC2998, a colorectal cancer cell line, harbors a heterozygous nonsense mutation of

TP53. There is no evidence of the second allele of *TP53* being altered by other mutations therefore the heterozygous truncating variant is tentatively oncogenic in HCC2998. Two MSI+ cell lines KM12 and MOLT-4 harbor heterozygous truncating variants of *TP53* (Figure 3-7). These variants are considered tentatively oncogenic because the truncating variants are heterozygous. In addition the nature of the variants is consistent with the presence of microsatellite instability (MSI+) in those cell lines.

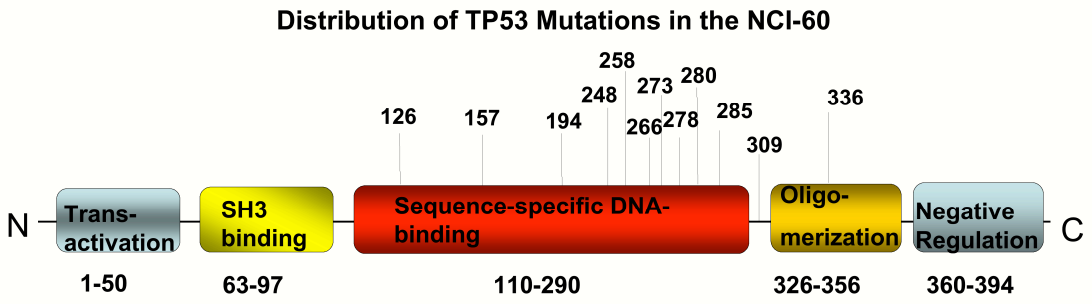

Figure 3-8: Distribution of TP53 mutations identified in the NCI-60 cell lines. A schematic of the TP53 protein, domains and location of the amino acid positions of the missense, nonsense, and frame-shift mutations identified in the NCI-60 cell lines.

3.3.15 BRCA2

Four variants of *BRCA2* were identified in three cell lines of the NCI-60 (Table 3-16). *BRCA2* is a tumor suppressor gene in breast cancer and the mutations of *BRCA2* are characterized by frame-shift mutations resulting in a truncated protein product. All of the mutations of *BRCA2* were identified in microsatellite unstable colon carcinoma cell lines (Figure 3-9).

Cell line	Tumor type	*BRCA2* variants identified	Type
HCT-116	Colon carcinoma	Het c.8021_8022insA p.I2675fsX6	TOV
HCT-15	Colon carcinoma	Het c.3599_3600delGT p.C1200fsX1, Het c.5351delA p.N1784fsX7	TOV
KM12	Colon carcinoma	Het c.5351delA p.N1784fsX7	TOV

Table 3-16. Variants of BRCA2 identified in the NCI-60 cell lines.

***BRCA2* cDNA sequence:**

A. *BRCA2* HCT-15 deletion of nucleotides 3599 and 3560:

gac**tgtaa**caaaagtgcttctggttatttaacagatgaaaatgaa
D C N K S A S G Y L T D E N E

B. *BRCA2* HCT-15 and KM12 deletion of nucleotide 5351:

a**a**cactagtttttccaaag**taa**tatccaatgtaaaagatgcaaat
N T S F S K V I S N V K D A N

C. *BRCA2* HCT-116 insertion between nucleotides 8021 and 8022:

gctataaaaaa**^a**gataatggaaaggga**tga**cacagctgcaaaaca
A I K K I M E R D D T A A K T

Figure 3-9: Illustration of nucleotide sequence changes in BRCA2 of the colorectal cancer cell lines HCT-15, KM12, and HCT-116. A) HCT-15 a microsatellite unstable (MSI+) colorectal cancer cell line harbors a deletion of a guanine and thymine at positions 3599 and 3560 (bold red). This causes a frame-shift resulting in a premature stop (taa) codon (bold black). B) HCT-15 and KM12, both microsatellite unstable (MSI+) colorectal cell lines harbor a deletion of an adenine at position 5351 (bold red). This causes a frame-shift resulting in a premature stop (taa) codon (bold black). C) HCT-116 an MSI+

colorectal cancer cell line harbors an insertion of an adenine between nucleotide positions 8021 and 8022 within a stretch of six adenine nucleotides (grey highlight). This causes a frame-shift resulting in a premature stop (tga) codon (bold black).

BRCA2 has not been reported as a cancer gene (either a susceptibility gene or somatically mutated) in colon cancer and somatic mutations of BRCA2 are generally very rare (COSMIC). Therefore, it is likely that the presence of heterozygous truncating variants of *BRCA2* in HCT-15, KM12, and HCT-116 are due to the microsatellite instability of the cell lines and do not contribute to tumorigenesis.

3.3.16 EGFR

EGFR is a dominantly acting cancer gene and its mutations are characterized by missense amino acid substitutions, and in frame insertions / deletions. Two mutations of *EGFR* were identified in each of two cell lines of the NCI-60 at codons 751 and 753 (Table 3-17). The mutations occurred in a myeloma cell line and a melanoma cell line. EGFR kinase domain mutations occur most frequently in lung cancers (COSMIC). It is rare to find *EGFR* mutations in myelomas and melanomas (COSMIC). Codons 751 and 753 are subject to missense substitutions and are involved in complex deletion/substitution mutations in lung adenocarcinoma (COSMIC, Figure 3-10). A complex somatic in-frame deletion mutation of *EGFR* involving codons 751 to 759 has been reported to occur in a primary lung cancer sample (Kang et al. 2007). Overall, therefore, their position provides evidence in favour of

the *EGFR* variants identified in the NCI-60 myeloma and melanoma cell lines being oncogenic. Further investigation of these lines for sensitivity to EGFR inhibitors and the potential role of *EGFR* mutations in a subset of melanoma and myeloma is, however, warranted.

Cell line	Tumor type	*EGFR* variants identified	Type
RPMI-8226	Myeloma	Het c.2252C>T p.T751I	LOM
SK-MEL-28	Melanoma	Hom c.2257C>T p.P753S	LOM

Table 3-17. Mutations of EGFR identified in the NCI-60 cell lines.

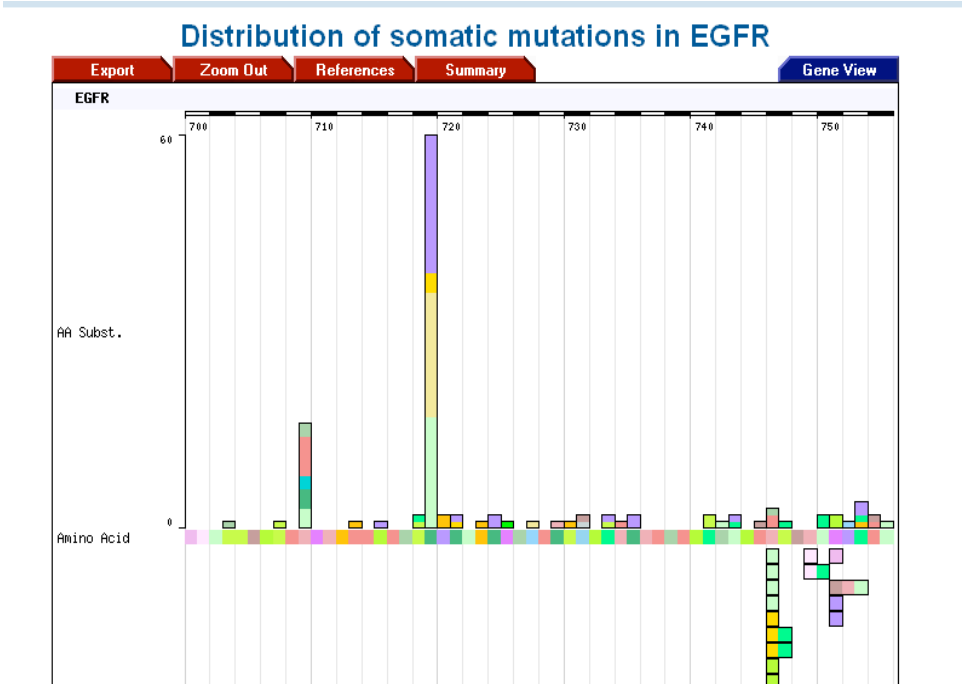

Figure 3-10: Distribution of somatic mutations in EGFR. A diagram illustrating the frequency of mutations in the kinase domain between codons 700 and 755. Adapted from COSMIC (http://www.sanger.ac.uk/perl/genetics/CGP/cosmic).

3.3.17 ERBB2

One mutation of *ERBB2* was identified in two out of seven ovarian carcinoma cell lines of the NCI-60. Mutations of *ERBB2* occur in 5% of ovarian cancers (COSMIC). *ERBB2* is a dominantly acting cancer gene and the mutation of *ERBB2* is characterized by a missense amino acid substitution at Glycine 776. Glycine 776 and the adjacent valine 777 are somatically mutated infrequently in gastric, lung and colon cancers (Lee et al. 2006, Stephens et al. 2004).

The two ovarian cancer cell lines have identical *ERBB2* mutations. I have found that OVCAR-8 and NCI/ADR-RES also have identical *TP53* mutations and share 99% genotype similarity across 10,000 SNPs (http://www.sanger.ac.uk/genetics/CGP/Genotyping/nci60.shtml). NCI/ADR-RES was once thought to be an adriamycin resistant form of MCF7 breast cancer cell line. However, gene expression analysis revealed that NCI/ADR-RES clustered more tightly with ovarian cancers (Ross et al. 2000). The identical sequence mutations and genotype similarity strongly suggest that NCI/ADR-RES is an ovarian cancer cell line identical to OVCAR-8.

Cell line	Tumor type	*ERBB2* variants identified	Type
OVCAR-8	Ovarian carcinoma	Het c.2327G>T p.G776V	LOM
NCI/ADR-RES	Ovarian carcinoma	Het c.2327G>T p.G776V	LOM

Table 3-18. Mutations of ERBB2 identified in the NCI-60 cell lines.

3.3.18 FLT3

One mutation of *FLT3* was identified in an acute lymphoblastic leukemia of the NCI-60 (Table 3-19). *FLT3* is a dominantly acting cancer gene and internal tandem duplications and point mutations of *FLT3* are frequent in acute myelogenous leukemia (COSMIC). The mutation of *FLT3* is characterized by a single missense amino acid substitution at codon 627. Alanine 627 is just adjacent to the G-loop ATP binding motif within the kinase domain and is very highly conserved.

Cell line	Tumor type	*FLT3* variant identified	Type
CCRF-CEM	Acute lymphoblastic leukemia	Het c.1879G>A p.A627T	TOV

Table 3-19. Variant of FLT3 identified in the NCI-60 cell lines.

3.3.19 PDGFRA

One donor splice site mutation at exon 10 of *PDGFRA* was identified in a chronic myeloid leukemia cell line of the NCI-60 (Table 3-20). *PDGFRA* is a dominantly acting cancer gene predominantly mutated by missense amino acid substitutions in Gastrointestinal stromal tumors (GISTs). Exon 10 of *PDGFRA* is homologous to *KIT* exon 9, which is the second most common site of *KIT* mutations in GISTs (Corless et al. 2005). *PDGFRA* is implicated in some cases of CML by the presence of the *BCR-PDGFRA* fusion gene with breakpoints usually within exon 12 of *PDGFRA*.

Cell line	Tumor type	*PDGFRA* variant identified	Type
K-562	Chronic myeloid leukemia	Het exon 10 +1 G>A p.?	TOV

Table 3-20. Variant of PDGFRA identified in the NCI-60 cell lines.

The consequences of the donor splice site variant in PDGFRA are unclear. Prediction programs suggest that, as a result of the splice site mutation, the donor splice site at position 44452 in *PDGFRA* would no longer be used. The next predicted donor splice site is position 45347, corresponding to the exon 11 donor splice site (Figure 3-11). Therefore the G>A nucleotide change at the donor splice site of exon 10 could result in the loss of splicing and introduction of a premature stop codon within the intron 10-11 sequence (Figure 3-12). This would yield a probably non-functional truncated protein (Figure 3-14) lacking most of the transmembrane domain and all of the kinase domain of PDGFRA (Figure 3-13). It is also possible, however, that the splice variant might result in exon 10 skipping, which would result in an in-frame change that could potentially be activating.

NetGene2 splice site prediction output:

A. *PDGFRA*_wt

Pos 5'->3'	strand	confidence	5' exon intron 3'
43174	+	0.93	GGGACAGACG^GTGAGGTGCA
44452	+	0.95	GTGGCTCCCA^**G**TGAGTTCCT
45347	+	0.71	TTGGAAACAG^GTAGATATTT

NetGene2 splice site prediction output:

B. *PDGFRA*_mut

Pos 5'->3'	strand	confidence	5' exon intron 3'
43174	+	0.93	GGGACAGACG^GTGAGGTGCA
45347	+	0.71	TTGGAAACAG^GTAGATATTT

Figure 3-11: Predicted splice sites in PDGFRA wild type and mutant (exon 10 +1 G>A) gene. A) Output of the predicted splice sites of PDGFRA and

confidence scores. The splice donor guanine nucleotide (bold red) is changed to an adenine in K562 chronic myeloid leukemia cell line. B) Output of the predicted splice sites of mutant (exon 10 +1 G>A) PDGFRA. Output from NetGene2 splice site prediction web server (http://www.cbs.dtu.dk/services/NetGene2/).

Exon 10 and intron 10-11 of PDGFRA:

ATGTAATAATGAAACTTCCTGGACTATTTTGGCCAACAATGTCTCAAACATC
ATCACGGAGATCCACTCCCGAGACAGGAGTACCGTGGAGGGCCGTGTGA
CTTTCGCCAAAGTGGAGGAGACCATCGCCGTGCGATGCCTGGCTAAGAAT
CTCCTTGGAGCTGAGAACCGAGAGCTGAAGCTGGTGGCTCCCA**g**tgagttcctc
aacagtcaggacaactcatcagctgagccgcatctgccccaggcggaactt**tga**atcccag

Figure 3-12: Illustration of PDGFRA exon 10 and intron 10-11 where the exon 10 + 1 G>A (bold red) mutation occurs. This mutation causes loss of the splice site and continues the reading frame into intron 10-11. The result is coding of a premature stop (tga) codon (bold black) within intron 10-11, resulting in a truncated protein product. Adapted from Ensembl genome browser (http://www.ensembl.org).

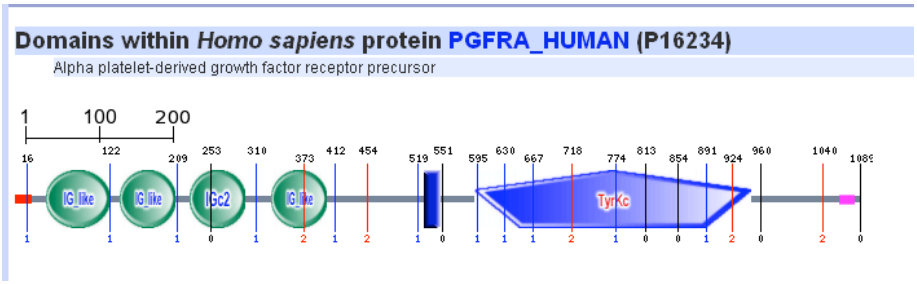

Figure 3-13: Illustration of the wild type PDGFRA protein encoding 1089 amino acids. Adapted from SMART (http://smart.embl-heidelberg.de/).

Figure 3-14: Illustration of the mutant (exon 10 +1 G>A) PDGFRA protein encoding 540 amino acids. Adapted from SMART (http://smart.embl-heidelberg.de/).

Table 3-21. Compilation of all mutations/variants identified in each of the NCI-60 cell lines

Cell Line	Tissue type	Variants Identified
786-0	Renal cell carcinoma	***CDKN2A*** Hom c.1_150 del 150, p.? LOM; ***PTEN*** Hom c.445C>T, p.Q149* LOM; ***TP53*** het c.832C>G, P278A c.A560-2A>G, p.? LOM; ***VHL*** Hom c.311delG p.G105fs*55 LOM
A498	Renal cell carcinoma	***CDKN2A*** Hom c.1_471 del 471, p.? LOM; ***VHL*** Hom c.426_429delTGAC p.G144fs*14 LOM
A549	Lung carcinoma	***CDKN2A*** Hom c.1_471 del 471, p.? LOM; ***KRAS*** Hom c.34G>A, p.G12S LOM; ***STK11*** Hom c.109C>T, p.Q37* LOM
ACHN	Renal cell carcinoma	***CDKN2A*** Hom c.1_471 del 471, p.? LOM
BT-549	Breast carcinoma	***RB1*** Hom c.265_607 del 343, p.? LOM; ***TP53*** Hom c.747G>C, p.R249S LOM
CAKI-1	Renal cell carcinoma	***CDKN2A*** Hom c.1_471 del 471, p.? LOM
CCRF-CEM	Acute lymphoblastic leukemia	***CDKN2A*** Hom c.1_471 del 471, p.? LOM; ***KRAS*** Het c.35G>A, p.G12D LOM; ***PTEN*** Hom c. del 80-492, p.? LOM; ***TP53*** Het c.743G>A, p.R248Q c.524G>A, p.R175H LOM; ***FLT3*** Het c.1879G>A p.A627T TOV
COLO-205	Colorectal cancer	***APC*** Hom c.4666_4667insA p.T1556fs*3 LOM; ***BRAF*** Het c.1799T>A, p.V600E LOM; ***SMAD4*** Hom del exon1-6 LOM; ***TP53*** Hom c.308_333>TA, p.Y103fs*37 LOM
DU-145	Prostate carcinoma	***CDKN2A*** Hom c.250G>T p.D84Y LOM; ***RB1*** Hom c.2143A>T, p.K715* LOM; ***STK11*** Hom c.532_536delAAGCC p.K178fs*86 LOM; ***TP53*** Het c.820G>T, p.V274F LOM
EKVX	Lung carcinoma	***TP53*** Hom c.609_610GG>TT, p.E204* LOM
HCC2998	Colorectal carcinoma	***APC*** Het c.1994T>A, p.L665* c.4348C>T, R1450* LOM; ***KRAS*** Het c.436G>A, p.A146T LOM; ***RB1*** Het c.411G>T, p.E137* TOV; ***TP53*** Het c.637C>T, p.R213* TOV

HCT-116	Colorectal carcinoma	***CDKN2A*** Het c.68_69insG p.R24fsX20 ***CDKN2A*** Het c.220delG p.E74fsX15 (p14) LOM; ***KRAS*** Het c.38G>A, p.G13D LOM; ***PIK3CA*** Het c.3140A>G, p.H1047R LOM; ***BRCA2*** Het c.8021_8022insA p.I2675fs*6 TOV
HCT-15	Colorectal carcinoma	***APC*** Het c.6496C>T,p.R2166* ***APC*** Hom c.4248delC p.I1417fs*2 LOM; ***KRAS*** Het c.38G>A, p.G13D LOM; ***PIK3CA*** Het c. 1633G>A p. E545K LOM; ***TP53*** Het c.C1101-2A>C, p.? C722T, S241F LOM; ***BRCA2*** Het c.3599_3600delGT p.C1200fs*1 ***BRCA2*** Het c.5351delA p.N1784fs*7 TOV
HL-60	Acute myeloid leukemia	***CDKN2A*** Hom c.238C>T p.R80X LOM; ***NRAS*** Het c.182A>T, p.Q61L LOM; ***TP53*** Hom deletion LOM
HOP62	Lung carcinoma	***CDKN2A*** Hom c.1_471 del 471, p.? LOM; ***KRAS*** Het c.34G>T, p.G12C LOM; ***TP53*** Hom c.G673-2A>G, p.? LOM
HOP-92	Lung carcinoma	***CDKN2A*** Hom c.1_471 del 471, p.? LOM; ***TP53*** Hom c.524G>T, p.R175L LOM
Hs-578-T	Breast carcinoma	***CDKN2A*** Hom c.1_471 del 471, p.? LOM; ***HRAS*** Het c.35G>A p.G12D LOM; ***TP53*** Hom c.469G>T, p.V157F LOM
HT-29	Colorectal carcinoma	***APC*** Het c.2557G>T p.E853X ***APC*** Het c.4666_4667insA p.T1556fs*3 LOM; ***BRAF*** Het c.1799T>A, p.V600E LOM; ***SMAD4*** Hom c.931C>T, p.Q311* LOM; ***PIK3CA*** Het c.1345C>A p.P449T LOM; ***TP53*** Hom c.818G>A, p.R273H LOM
IGROV-1	Ovarian carcinoma	***TP53*** Het c.377A>G, p.Y126C LOM; ***BRCA1*** Het c.1961delA p.K654fs*47 TOV; ***SMAD4*** Het c.692delG p.G231fsX10 TOV; ***PTEN*** Het c.955_958delACTT p.T319fsX1 TOV
K-562	Chronic myeloid leukemia	***CDKN2A*** Hom c.1_471 del 471, p.? LOM; ***TP53*** Hom c.406_407insC p.Q136fs*13 LOM; ***PDGFRA*** Het Exon 10 +1 G>A p.? TOV
KM12	Colorectal carcinoma	***PTEN*** Het c.385G>T, p.G129* ***PTEN*** Het c.800 del A, p.K267fs*9 LOM; ***APC*** Het c.5454_5455 ins A, p.N1819fs*7 TOV; ***TP53*** Het 215 del G, V122* TOV; ***BRCA2*** Het c.5351delA p.N1784fs*7 TOV
LOXIMVI	Melanoma	***BRAF*** Het c.1799T>A, p.V600E LOM; ***CDKN2A*** Hom c.1_471 del 471, p.? LOM
M14/MDA-MB-435	Melanoma	***BRAF*** Het c.1799T>A, p.V600E LOM; ***CDKN2A*** Het c.150+2 T>C p.? ***CDKN2A*** Het c. 456-+24 AGgtgaggactgatgatctgagaattt >C p.? LOM; ***TP53*** Het c.797G>A, p.G266E LOM
MALME-3M	Melanoma	***BRAF*** Het c.1799T>A, p.V600E LOM; ***CDKN2A*** Hom c.1_471 del 471, p.? LOM
MCF7	Breast carcinoma	***CDKN2A*** Hom c.1_471 del 471, p.? LOM; ***PIK3CA*** Het c.1633G>A p.E545K LOM
MDA-MB-231	Breast carcinoma	***BRAF*** Het c.1391G>T, p.G464V LOM; ***CDKN2A*** Hom c.1_471 del 471, p.?; ***KRAS*** Het c.38G>A p.G13D LOM;

		TP53 Hom c.839G>A, p.R280K LOM
MOLT-4	Acute lymphoblastic leukemia	***CDKN2A*** Hom c.1_471 del 471, p.? LOM; ***NRAS*** Het c.34G>T p.G12C LOM; ***PTEN*** Hom c.800delA p.K267fs*9 LOM; ***STK11*** Het c.640C>T p.Q214X TOV; ***TP53*** Het c.916C>T, p.R306* TOV
NCI-H226	Lung squamous cell carcinoma	***CDKN2A*** Hom c.1_150 del 150, p.? LOM
NCI-H23	Lung adenocarcinoma	***KRAS*** Het c.34G>T, p.G12C LOM; ***STK11*** Hom c.996T>A, p.W332X LOM; ***TP53*** Hom c.738G>C, p.M246I LOM
NCI-H322M	Lung bronchoalveolar carcinoma	***TP53*** Hom c.743G>T, p.R248L LOM
NCI-H460	Lung large cell carcinoma	***CDKN2A*** Hom del 1_457 del 457, p.? LOM; ***KRAS*** Hom c.183A>T, p.Q61H LOM; ***PIK3CA*** Het c.1633G>A p.E545K LOM
NCI-H522	Lung adenocarcinoma	***TP53*** Hom 572delC, p.P191fs*56 LOM
OVCAR3	Ovarian carcinoma	***TP53*** Hom c.743G>A, p.R248Q LOM
OVCAR-4	Ovarian carcinoma	***TP53*** Hom c.388C>G, p.L130V LOM
OVCAR-5	Ovarian carcinoma	***CDKN2A*** Hom c.1_471 del 471, p.? LOM; ***KRAS*** Hom c.35G>T, p.G12V LOM
OVCAR-8/ NCIADR-RES	Ovarian carcinoma	***TP53*** Hom c.376-1G>A, p.? LOM; ***ERBB2*** Het c.2327G>T p.G776V LOM
PC-3	Prostate carcinoma	***PTEN*** Hom c.165-1026 del 861, p.? LOM; ***TP53*** Hom c.414delC p.K139fs*31 LOM
RPMI-8226	Myeloma	***KRAS*** Het c.35G>C, p.G12A LOM; ***TP53*** Hom c.853G>A, p.E285K LOM; ***EGFR*** Het c.2252C>T p.T751I LOM
RXF393	Renal cell carcinoma	***CDKN2A*** Hom c.1_471 del 471, p.? LOM; ***PTEN*** Hom c.1_164 del 164, p.? LOM; ***TP53*** Hom c.524G>A, p.R175H LOM
SF-268	Glioma	***CDKN2A*** Hom c.1_471 del 471, p.? LOM; ***TP53*** Hom c.818G>A, p.R273H LOM
SF-295	Glioma	***CDKN2A*** Hom c.1_471 del 471, p.? LOM; ***PTEN*** Hom c.694C>T, p.R232* LOM; ***TP53*** Hom c.743G>A, p.R248Q LOM
SF539	Glioma	***RB1*** Hom c.346_349delACTT p.T116fs*8 LOM; ***TP53*** Hom c.1024delC p.R342fs*3 LOM; ***PTEN*** Hom c.1-1026 del 1026, p.? LOM
SK-MEL-2	Melanoma	***NRAS*** Hom c.182A>G, p.Q61R LOM; ***TP53*** Het c.733G>A, p.G245S LOM
SK-MEL-28	Melanoma	***BRAF*** Hom c.1799T>A, p.V600E LOM; ***TP53*** Hom c.435_436TG>GT, p.L145R LOM; ***EGFR*** Hom c.2257C>T p.P753S LOM
SK-MEL-5	Melanoma	***BRAF*** Het c.1799T>A, p.V600E LOM; ***CDKN2A*** Hom c.1_471 del 471, p.? LOM
SK-OV-3	Ovarian carcinoma	***CDKN2A*** Hom del 1_457 del 457, p.? LOM; ***PIK3CA*** Het c.3140A>G, p.H1047R LOM;

		TP53 Hom c. del267C p.S90fs*33 LOM; ***APC*** Het c.4666delA p.T1556fsX9 TOV
SN12C	Renal cell carcinoma	***TP53*** Hom c.1006G>T, p.E336* LOM
SNB-75	Glioma	***TP53*** Hom c.772G>A, p.E258K LOM
SR	Non Hodgkin lymphoma	***CDKN2A*** Hom c.1_471 del 471, p.? LOM
SW620	Colorectal carcinoma	***KRAS*** Hom c.35G>T, p.G12V LOM; ***TP53*** Hom c.818G>A, p.R273H C925T, p.P309S LOM
T47D	Breast carcinoma	***PIK3CA*** Het c.3140A>G, p.H1047R LOM; ***TP53*** Hom c.580C>T, p.L194F LOM
U251/SNB-19	Glioma	***CDKN2A*** Hom c.1_471 del 471, p.? LOM; ***PTEN*** Hom c.723_724insTT p.E241fs*15 LOM; ***TP53*** Hom c.818G>A, p.R273H LOM
UACC-257	Melanoma	***BRAF*** Het c.1799T>A, p.V600E LOM
UACC-62	Melanoma	***BRAF*** Hom c.1799T>A, p.V600E LOM; ***CDKN2A*** Hom c.1_471 del 471, p.? LOM; ***PTEN*** Hom c.741_742insA p.P248fs*5 LOM
UO-31	Renal cell carcinoma	***CDKN2A*** Hom c.1_471 del 471, p.? LOM

3.4 Discussion

A total of 141 LOMs and 15 TOVs were identified in the 24 cancer genes sequenced in the NCI-60 (Figure 3-15). Taking into account point mutations, small insertions/deletions and homozygous deletions, 16 of 24 cancer genes were found to have likely oncogenic mutations (LOM) in at least one cell line (*APC, BRAF, CDKN2A, CTNNB1, EGFR, ERBB2, HRAS, KRAS, NRAS, SMAD4, PIK3CA, PTEN, RB1, STK11, TP53, and VHL*), and 3 of the 24 cancer genes were found to have only tentative oncogenic variants (TOV) (*BRCA2*, *FLT3, PDGFRA*). I did not identify LOMs or TOVs in *BRCA1, KIT, MAP2K4, MET* and *RET*. By re-sequencing of cancer genes coupled with genotyping of 10,000 SNPs I found there are actually 57 independent cell lines as opposed to 60 represented in the NCI-60 (Ikediobi et al. 2006). Also, I found that several previously identified mutations of *TP53* are incorrect, due mostly to annotation errors.

TP53, the most commonly mutated gene in cancer, was altered in 64% (38/59) of the cell lines. *CDKN2A* single- or multiple-exon deletions/point mutations were observed in 56% (33/59) of the NCI-60 cell lines. Conversely likely oncogenic mutations were detected only once each in the *HRAS* and *CTNNB1* genes. The number of analyzed cancer genes with likely oncogenic mutations in any single cell line ranged from five in the microsatellite-stable colorectal cancer line HT-29 (*APC, BRAF, SMAD4, PIK3CA*, and *TP53*) to one in several lines (Figure 3-16).

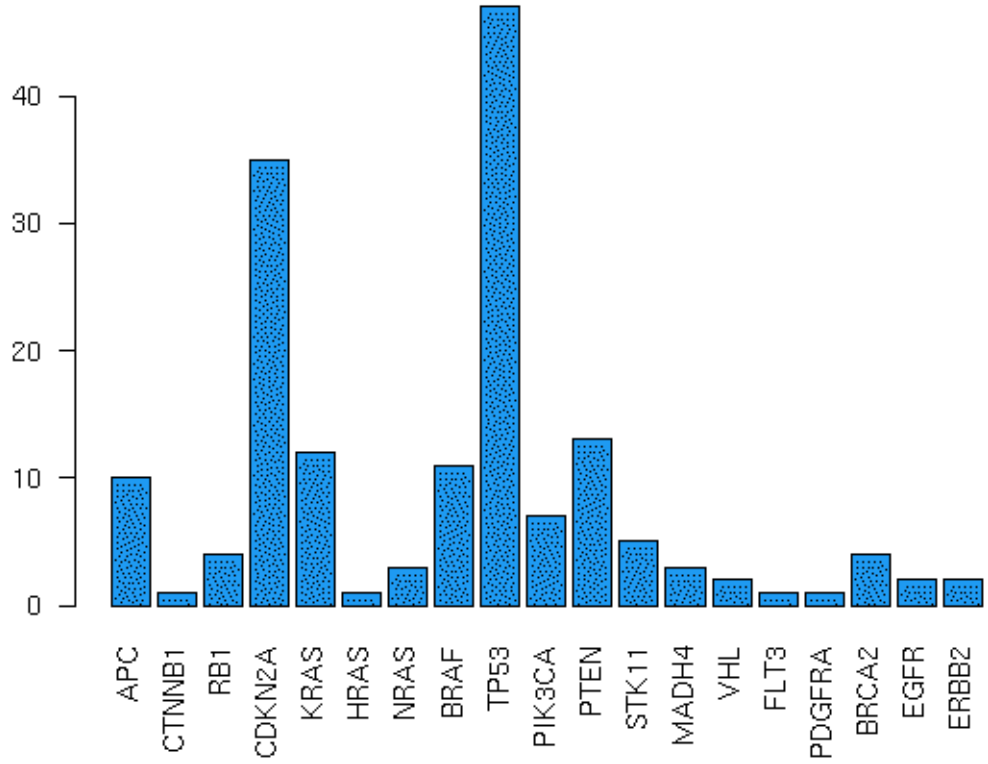

Figure 3-15: The total number of variants (LOMs and TOVs) identified in 19 of 24 cancer genes sequenced in the NCI-60 cell lines. The x-axis is labeled with each of the 19 genes in which I identified sequence variants. The y-axis represents the number of variants identified for each gene.

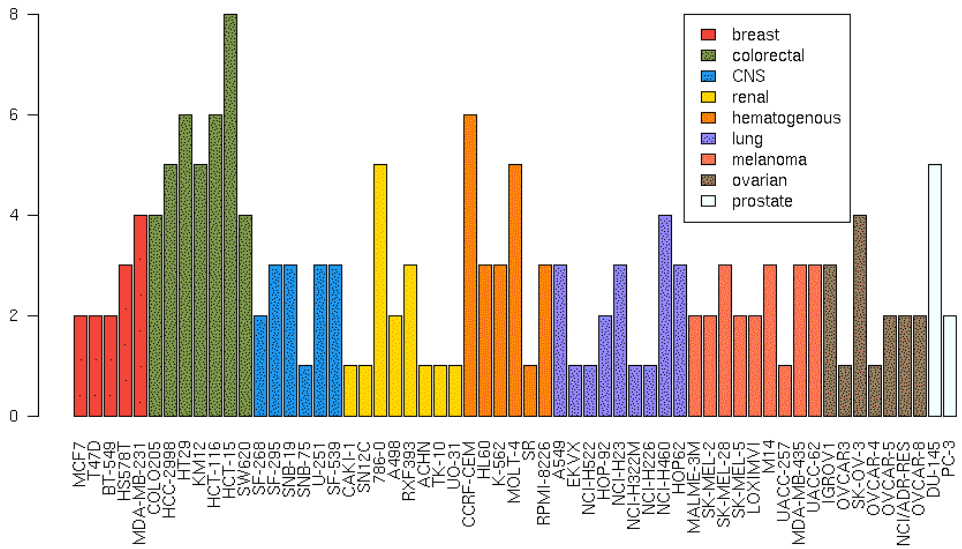

Figure 3-16: Total number of variants (LOMs and TOVs) identified in each of the 59 cell lines of the NCI-60. The cell lines are grouped according to tissue-of-origin.

A mutational screen of 518 kinase genes in 210 human cancers (274 Mb) has revealed that lung carcinomas have the highest prevalence of somatic mutations (4.21 per Mb), followed by gastric cancers (2.1 per Mb), ovarian cancers (1.85 per Mb), colorectal cancers (1.21 per Mb) and renal cancers (0.74 per Mb) (Greenman et al. 2007). Based on those frequencies of somatic mutations in human cancer I can extrapolate to that observed in the NCI-60 cell line panel. The highest prevalence of variants in the NCI-60 was observed in the colon cancer cell lines (575 per Mb), followed by lung cancer cell lines (287 per Mb), ovarian cancer cell lines (227 per Mb) and

renal cancer cell lines (227 per Mb). Overall, the number of variants identified in the NCI-60 cell lines is much higher than that observed in primary tumors. This higher mutational frequency is mainly due to the fact that I focused on the sequencing of cancer genes. In contrast, the kinase screen was not biased for genes frequently mutated in cancer. As well, cancer cell lines are known to accumulate more mutations during *in vitro* culture.

I observe that the colorectal carcinoma cell lines of the NCI-60 have more variants, an average of five variants per cell line, compared to the other tissue types represented in the NCI-60 panel (Figure 3-16). The numbers of cancers studied and mutations observed are small, but these may reflect real differences between cancer types. This variability in mutation prevalence could be because I have identified more of the cancer genes involved in colon cancer tumorigenesis than in other tissue types represented in the NCI-60 panel. There is also an overrepresentation of microsatellite unstable colorectal cancers in the NCI-60 panel which could also increase the mutation frequency, since MSI+ cancers generally have a high mutation prevalence. The differences may also reflect a tendency of other cancers to use other mutational mechanisms including amplification and gene rearrangement. If so, the results indicate that collectively more is known about some cancer types than others making the case for continued systematic sequencing of more cancer genomes to fully understand the contribution of DNA sequence changes to the cancer phenotype. It is however, also conceivable, that some cancers require more mutations than others during tumorigenesis.

It is of interest to consider coincident combinations of cancer gene mutations. In the NCI-60 cell line panel the most common combination of cancer gene mutations occurs with *CDKN2A* and *TP53* (Figure 3-17), followed by *TP53* and *PTEN*. *CDKN2A* and *TP53* are mutated each in more than 50% of the cell lines. Therefore there is a high likelihood of finding coincident mutations of both genes. One cell line, DU145, harbors coincident mutations of *CDKN2A* and *RB1* genes. It is rare to observe mutations of both *CDKN2A* and *RB1* in primary tumors because mutations of either affect the same pathway. However, the *CDKN2A* variant in DU145 is a missense amino acid substitution with unclear biological implications. In the same vein one cell line, MDA-MB-231, harbors coincident mutations of *RAS* and *BRAF.* I did not detect cell lines with coincident mutations of *PIK3CA* and *PTEN*, however three cell lines harbored coincident mutations of RAS and PIK3CA. Though interesting, the weakness of such an analysis in the NCI-60 panel, composed of 57 cell lines, is the limited power to detect frequencies of such combinations. It would be of interest to extend this analysis to a larger set of cancer samples.

The NCI-60 cell lines only include 9 tissues-of-origin cancer types. Of the cancer types it does represent, there are small numbers in each group. In terms of the genetics I observe some similar patterns to primary tumors, for example the presence of *TP53* mutations in over 50% of the cell lines. However, there are some aberrant mutation patterns such as the presence of *RAS* or *BRAF* mutations in breast cancer cell lines, the low frequency of *VHL* mutation in renal cell carcinomas, the high frequency of *CDKN2A* mutations in

all tissue types, overrepresentation of microsatellite unstable cell lines in the colorectal cancers, high frequency of *TP53* mutations in prostate and hematopoietic cancers, and high frequency of *PTEN* and *RB1* mutations in prostate carcinoma lines. This distorted representation of mutated genes compared to primary tumours could be attributable to selection bias for a subset of primary tumors that tend to grow well *in vitro*. Moreover, cell lines have had further opportunity to evolve in culture and may acquire more mutations compared to primary tumors. It could also, of course, be due to misattribution of tissue types. On the other hand, unlike primary tumors, cancer cell lines are not contaminated by stromal tissue, therefore mutation detection is of higher sensitivity and mutation counts in primary tumours may have been underestimated.

With respect to the utility of the NCI-60 for informing on commonly mutated cancer genes as drug targets, over 50% of the NCI-60 are *TP53* mutant, and the mutations cluster in the sequence specific binding domain of TP53. Whereas restoring TP53 pathway function in *TP53* wild-type cancer cells continues to be a focus of intensive drug development efforts (Klein et al. 2004), restoring TP53 function in cells with mutant *TP53* remains challenging.

Another pharmaceutical goal is the development of small molecule inhibitors that act as functional mimetics for tumor suppressor genes. Inhibitors of the CTNNB1:TCF protein complex, for example, appear to target the WNT pathway specifically (Lepourcelet et al. 2004). In the NCI-60, six of seven colon cancer cell lines (COLO-205, HCC2998, HCT-116, HCT-15, HT-

29, SW620) have likely oncogenic mutations in *APC* or *CTNNB1*. Therefore the NCI-60 panel of colon cancer cell lines could potentially be used to predict the efficacy of small molecule inhibitors of the CTNNB1:TCF complex.

There are a number of *RAS* mutants in the NCI-60 cell line panel. To date, direct inhibition of activated *RAS* and hence, its downstream effector has not been effective in cancer therapy. Downstream of RAS there are several *BRAF* mutations in the NCI-60. Recently, the *BRAF* mutant lines of the NCI-60 have been found to be sensitive to kinase inhibitors of the downstream *BRAF* effector/signalling target mitogen-activated protein/ extracellular signal-regulated kinase kinase (MEK) (Solit et al. 2006).

The mutations in *PIK3CA* and *PTEN* suggest the panel may be of value for analysis of compounds that target the phosphatidylinositol 3-kinase (PI3K) pathway. *PIK3CA*, a lipid kinase, is a clear target for therapeutic development (Whyte and Holbeck et al. 2006, Stephens et al. 2005).

The receptor tyrosine kinases are perhaps the most successfully exploited set of molecular targets in cancer to date. Several family members (*EGFR, ERBB2, FLT3, KIT, MET, and PDGFRA*) were included in the set of 24 genes assessed. A number of interesting variants were identified. Three likely oncogenic mutations were identified in *EGFR* and *ERBB2*. Two mutations in the kinase domain of EGFR were identified in a myeloma line and a melanoma line, and one mutation of ERBB2 was identified in an ovarian line. Fifteen tentative oncogenic variants were identified, including a splicing

and missense variant in the receptor tyrosine kinase genes *PDGFRA* and *FLT3,* respectively. The remainder of this class consisted of heterozygous frame-shift mutations in tumor suppressor genes found primarily in microsatellite-unstable lines (HCT-15, HCT-116, KM12, IGROV-1, MOLT-4, SKOV3) (http://www.sanger.ac.uk/genetics/CGP/MSI/cell_line_page.shtml).

It has also become clear that mutations in tumor suppressor genes are not always completely recessive (Payne and Kemp 2005). The phenotype of mutations in tumor suppressor genes can be due to reduction in gene dosage (haploinsufficiency) and/or may act in concert with other oncogenic or haploinsufficient events. The phenotype may also depend on the genetic background and tissue type. Mouse models of heterozygous *PTEN* mutant prostate cancer reveal that haploinsufficiency of *PTEN* promotes progression of prostate cancer (Kwabi-Addo et al. 2001). Also, haploinsufficiency of *MADH4* in mice has been shown to initiate gastric polyposis with loss of the wild type allele during the later stages of tumorigenesis (Xu et al. 2000). Inactivation of one allele of *STK11* in mouse leads to the formation of gastrointestinal polyps (Miyoshi et al. 2002). It is then also possible that for some of the tumor suppressor genes, such as *PTEN*, *MADH4*, and *STK11* haploinsufficiency may be enough to contribute to the human cancer phenotype (Alberici et al. 2006).

The NCI-60 cell line set has been mostly used for the screening of anticancer activity of thousands of natural and synthetic compounds. To date, several lead anticancer compounds have been identified and some are in

clinical use. However, for the purposes of molecularly targeted screens, the NCI-60 may not be the most suitable or representative set of cell lines. The NCI-60 cell lines were assembled in the early 1990s before the advent of high throughput molecular genetic profiling. Based on the mutation analysis of 24 cancer genes in the NCI-60, I find that the cell line panel does not fully represent the genetic diversity seen in primary tumors. The NCI-60 is the most characterized set of cancer cell lines. It currently serves as a rich public resource for interrogating relationships between cancer genetics and drug response. However, in the future, it may be necessary to reconsider the panel of cell lines for molecularly targeted *in vitro* drug screens.

APC	*APC*																	
CTNNB1	0	*CTNNB1*																
RB1	1	0	*RB1*															
CDKN2A	1	1	1	*CDKN2A*														
KRAS	3	1	1	7	*KRAS*													
HRAS	0	0	0	1	0	*HRAS*												
NRAS	0	0	0	2	0	0	*NRAS*											
BRAF	2	0	0	7	1	0	0	*BRAF*										
TP53	7	0	4	18	8	1	3	6	*TP53*									
PIK3CA	3	1	0	4	3	0	0	1	4	*PIK3CA*								
PTEN	1	0	1	8	1	0	1	1	11	0	*PTEN*							
STK11	0	0	1	4	3	0	1	0	3	1	1	*STK11*						
MADH4	2	0	0	0	0	0	0	2	3	1	1	0	*MADH4*					
VHL	0	0	0	2	0	0	0	0	1	0	1	0	0	*VHL*				
FLT3	0	0	0	1	0	0	0	0	1	0	1	0	0	0	*FLT3*			
PDGFRA	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	*PDGFRA*		
BRCA2	2	1	0	1	2	0	0	0	2	2	1	0	0	0	0	0	*BRCA2*	
EGFR	0	0	0	0	1	0	0	1	2	0	0	0	0	0	0	0	0	*EGFR*
ERBB2	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0

Figure 3-17: Display of the number of mutation combinations identified in the NCI-60 cell line panel. Values highlighted in red represent the most common mutation combinations represented in the NCI-60.