Combinatorial patterns of somatic gene mutations in cancer

Chen-Hsiang Yeang,*,[1] Frank McCormick,[†] and Arnold Levine*

*Simons Center for Systems Biology, Institute for Advanced Study, Princeton, New Jersey, USA; and the [†]Helen Diller Family Comprehensive Cancer Center and Cancer Research Institute, University of California, San Francisco, California, USA

ABSTRACT **Cancer is a complex process in which the abnormalities of many genes appear to be involved. The combinatorial patterns of gene mutations may reveal the functional relations between genes and pathways in tumorigenesis as well as identify targets for treatment. We examined the patterns of somatic mutations of cancers from Catalog of Somatic Mutations in Cancer (COSMIC), a large-scale database curated by the Wellcome Trust Sanger Institute. The frequently mutated genes are well-known oncogenes and tumor suppressors that are involved in generic processes of cell-cycle control, signal transduction, and stress responses. These "signatures" of gene mutations are heterogeneous when the cancers from different tissues are compared. Mutations in genes functioning in different pathways can occur in the same cancer (*i.e.*, co-occur), whereas those in genes functioning in the same pathway are rarely mutated in the same sample. This observation supports the view of tumorigenesis as derived from a process like Darwinian evolution. However, certain combinatorial mutational patterns violate these simple rules and demonstrate tissue-specific variations. For instance, mutations of genes in the Ras and Wnt pathways tend to co-occur in the large intestine but are mutually exclusive in cancers of the pancreas. The relationships between mutations in different samples of a cancer can also reveal the temporal orders of mutational events. In addition, the observed mutational patterns suggest candidates of new cosequencing targets that can either reveal novel patterns or validate the predictions deduced from existing patterns. These combinatorial mutational patterns provide guiding information for the ongoing cancer genome projects.—Yeang, C-H., McCormick, F., Levine, A. Combinatorial patterns of somatic gene mutations in cancer. *FASEB J.* 22, 2605–2622 (2008)**

Key Words: cell cycle control · Ras pathway · Wnt pathway · P53 pathway · Igf · Akt pathway · TGFβ pathway

Decades of studies have identified a large number of oncogenes and tumor suppressors and have placed them in the signaling and regulatory pathways. In most cases, however, the studies have focused on single genes or single pathways. The challenge is to understand how various cellular and physiological processes are coordinatedly altered during the progression of cancer. This question has led to the projects to sequence many or all of the genes in a large number of cancers and examine the collective mutations that occur as the tumor evolves (*e.g.*, the Cancer Genome Project initiated by the Wellcome Trust Sanger Institute, ref. 1; the Cancer Genome Atlas project launched by the U.S. National Cancer Institute, ref. 2). However, despite the fact that sequencing technologies have become cheaper and faster, large amounts of resources are still required to sequence many samples of all the common tumors, and the completion of these ambitious projects may still be years ahead.

While the completion of these large-scale sequencing/resequencing projects will provide unparalleled quantities of information about genetic and molecular alterations of cancer, the data accumulated from decades of published work can already reveal the recurrent mutational patterns and crosstalk of major cancer pathways. In this review, we present a useful and unique analysis of the somatic mutation data from thousands of previous studies and identify the recurrent combinatorial mutational patterns in 45 different tissue types. The results demonstrate the heterogeneity of combinatorial patterns in different tissues, confirm the functional relations of genes in the pathways, indicate the differential couplings between pathways in different tissues, and reveal the temporal orders of mutational events. Furthermore, using the information of the combinatorial patterns we suggest candidates for new sequencing projects that can either reveal novel patterns or validate predictions deduced from existing patterns.

DATABASE OF SOMATIC GENE MUTATIONS IN CANCER

The Catalog of Somatic Mutations in Cancer (COSMIC) is a large database of cancer somatic gene mutations curated by the Wellcome Trust Sanger Institute (3). It extracts from a large number of publications the mutational records of ~3000 genes in various types of

[1] Correspondence: Institute for Advanced Study, 1 Einstein Drive, Princeton, NJ 08540, USA. E-mail: chyeang@ias.edu

doi: 10.1096/fj.08-108985

primary tumor tissues or cancer cell lines. The version of May 2007 contains 218,323 cancer samples and 514,020 records extracted from 4138 studies. Currently, COSMIC collects only small mutations in protein-coding regions, including point mutations, small insertions and deletions, frame shifts, and unspecified mutations in the sources.

As most studies were undertaken before the post-genomic era, when large-scale sequencing was either unattenable or nonaffordable, the majority of the 218,323 samples probe only one (89.71%) or a few genes. As expected, most studies target only a small set of well-known oncogenes and tumor suppressors, such as TP53, KRAS, and APC. An exception is the 785 cell line samples where large-scale screenings of 3303 genes were performed by the Sanger Institute (Cancer Cell Line Project; ref. 4).

The sparseness of genes probed in most samples explains the low frequency of mutations: only 19% (41,475 of 218,323) of the samples contain at least one detected mutation, and only 9.04% (46,491 of 514,020) of the records are mutations. However, even common oncogenes or tumor suppressors are not mutated in more than half of the samples in which they are probed. TP53 mutation, for instance, comprises only 46% (419 of 902) of the samples probed for this gene.

MUTATIONAL SIGNATURES ARE HETEROGENEOUS IN DIFFERENT TISSUES

Among the 3303 genes examined in the database, only 44 genes are mutated in more than 10 samples within any of the 45 tissue types tested. **Table 1** lists the 44 genes and their mutational frequencies in 17 selected tissue types. (The mutational frequency table of all 45 tissue types is reported in the Supplemental Material.) Nearly all genes are known oncogenes or tumor suppressors in major signaling or regulatory pathways of cancer. Many of them are involved in cell growth and proliferation (Ras pathway: KRAS, NRAS, HRAS, BRAF; IGF-AKT pathway: PTEN, PIK3CA; Wnt pathway: APC, β-catenin; TGF-β pathway: SMAD4), cell cycle control (TP16, RB1), and stress response and apoptosis (TP53).

The mutational signatures of those genes are heterogeneous in different tissues. Some genes, such as TP53 and TP16, are frequently mutated across most tissues. Other common oncogenes and tumor suppressors, by contrast, exhibit clear tissue specificity. For instance, APC mutations are found primarily in gastrointestinal tissues such as pancreas (24 of 152 samples), stomach (113 of 606 samples), large intestine (1192 of 3602 samples), and small intestine (31 of 211 samples). They are rare in lung adenocarcinoma (2 of 145 samples) and lung small cell carcinoma (1 of 144 samples). By contrast, RB1 mutations appear primarily in lung small cell carcinoma (52 of 116 samples), central nervous system (17 of 249 samples), urinary tract (18 of 54 samples), and eye tissue (76 of 164 samples). They are rare in large intestine tissue (1 of 78 samples).

Other genes are mutated only in a specific set of cancers. For example, FLT3 encodes a receptor tyrosine kinase that regulates some events in hematopoiesis (5–7) and is frequently mutated in acute myeloid leukemia (AML) (3108 of 13,468 samples). WT1 encodes a zinc-finger transcription factor involved in the development of the urogenital system (8) and is mutated primarily in kidney (Wilms tumor, 86 of 975 samples). EGFR encodes an epidermal growth factor receptor (9–11) and is mutated commonly in lung adenocarcinoma (713 of 3783 samples) and non small cell carcinoma (1303 of 3679 samples). Note that mutations in hematopoietic/lymphoid (*e.g.*, FLT3) and lung (*e.g.*, EGFR) cancers are often specific for certain subtypes of a tumor (*e.g.*, FLT3 in acute myeloid leukemia and EGFR in adenocarcinoma/non-small cell carcinoma).

SOME DIFFERENCES BETWEEN TUMOR AND CELL LINE MUTATIONAL FREQUENCIES ARE STATISTICALLY SIGNIFICANT

Since most cancer drug targets were commonly identified from cell lines, it is important to know whether the mutational signatures between primary tumors and cell lines derived from those types of tumors are different. COSMIC contains the data extracted from 785 cell line samples. The remaining 217,538 samples are generated from small-scale studies, which are primarily (but not exclusively) tumor tissues. **Table 2** shows the mutational frequencies in cell lines and tumors. We employed the gene-tissue entries where the genes were probed in more than 50 samples in both cell lines and tumors. For each gene-tissue combination, we evaluate the P value of the null hypothesis that the mutational frequencies of the cell line and tumor data are identical, and report the significant patterns in **Table 3** (also see the Supplemental Material). Six mutational signatures have significant differences ($P<0.001$). The disparate differences between cell lines and tumors are often confounded by the unbalanced sample sizes between the two subsets. More samples of cell lines or tumors are needed in order to understand whether these differences are truly meaningful.

TYPES OF SOMATIC MUTATIONS ARE CONSISTENT WITH THE FUNCTIONS OF GENES

We classified genes according to their functions as oncogenes or tumor suppressors and categorized the mutations in COSMIC into 5 classes: nonsynonymous point mutations, insertions (including tandem duplication and gene fusion) and deletions of coding sequences that remain in the reading frame, insertions or deletions that shift the reading frame (frame shifts),

and others (including synonymous point mutations and the records in which the positions of alterations are not specified). **Table 4** shows the occurrences of each type of mutation in each gene among all tissues. The refined version of Table 4 subdivided by tissue type and cell line/non-cell line data is reported in the Supplemental Material. Frame shifts occur almost exclusively in tumor suppressors, as they disrupt the proteins and their tumor-suppressing functions. Point mutations are common in both oncogenes and tumor suppressors, since they may either enhance or degrade the function of a protein. Insertions that result from translocations occur primarily in oncogenes (FLT3, 2915; KIT, 173; EGFR, 46; ERBB2, 32), except tumor-suppressor CEBPA (23 insertions). Many insertions are tandem duplications of protein domains (gene amplifications) (*e.g.*, in FLT3, ref. 12, and EGFR, ref. 13), which may enhance the functions of oncogenes. By contrast, deletions have been observed in both oncogenes (*e.g.*, EGFR, 728; KIT, 406; β-catenin, 182; PDGFRA, 46) and tumor suppressors (*e.g.*, TP16, 964; PTEN, 48; CEBPA, 11). Most deletions are detrimental to protein functions and thus should be observed in tumor suppressors. However, deletion of the extracellular domains of receptors (*e.g.*, EGFR; ref. 14) may cause a ligand-independent firing and promote cell growth. Moreover, deletions of the amino acid residues near the NH_2-terminal of β-catenin stabilize the protein against degradation and thus enhance its function (15).

COMBINATORIAL MUTATIONAL PATTERNS REVEAL THE FUNCTIONAL RELATIONS OF GENES IN MAJOR CANCER PATHWAYS

Combinatorial patterns of multiple gene mutations are closely linked to the functional relations of the genes in various processes of cancer formation. The progression of somatic mutations in cancer can be viewed as a Darwinian evolutionary process (16, 17). Tumorigenesis is a consequence of a series of mutations accumulated over years. Mutations of two genes participating in the same pathway or process rarely confer a significant selective advantage compared to the single mutation, since the functional consequences of single and double mutations are similar. By contrast, functional consequences of mutations of multiple genes that participate in different pathways or functions may be additive or even synergistic in conferring an advantage to the tumor. Therefore, we would expect to observe a tendency of mutually exclusive mutations of genes in the same pathway and the tendency of co-occurring mutations of genes that populated distinct pathways.

We categorize the mutations of multiple genes into two types of combinatorial patterns: co-occurrence of mutations in multiple ($\geq$2) genes and mutual exclusion of mutations in two genes. Mutual exclusion of more than two genes can be derived from pairwise relations because mutual exclusion relations are transitive. We define a test statistic for combinatorial patterns as the likelihood ratio (LR) between the empirical frequency of co-occurrence and the expected frequency according to the best simpler model that fits the data (see Supplemental Material). For two genes g_1 and g_2, this score is reduced to $P(g_1,g_2 \text{ mutated})/P(g_1 \text{ mutated})P(g_2 \text{ mutated})$. High scores indicate co-occurrence and low scores suggest mutual exclusion. The cutoff values of the score separating co-occurrence and mutual exclusion are determined by the background distribution of the scores (see Supplemental Material). To eliminate the effect of a small sample size we also evaluate the P value of the likelihood ratio (see Supplemental Material) and report only the combinatorial patterns with values of $P \leq 0.05$.

We obtain 105 significant combinatorial mutational patterns. The false discovery rate of permutation tests on the data is <0.01. **Table 5** shows the frequencies of significant combinatorial mutational patterns in different tissues. These combinatorial patterns cover genes in 6 major pathways relevant to cancer: cell cycle control, stress response, Ras, insulin growth factor (IGF-AKT), Wnt, and TGF-β signaling pathways. The majority of the combinatorial patterns conform with the simple hypotheses of pathways. **Table 6** counts the combinatorial patterns covering genes in each pair of pathways. Most co-occurring patterns contain genes in different pathways (off-diagonal entries in the top section of the table), whereas most mutually exclusive pairs are the genes in the same pathways (diagonal entries in the bottom section of the table). Furthermore, many combinatorial patterns appear in multiple tissues and multiple studies, suggesting they are not the artifacts from the contamination of tissue samples or special cases of certain tumor types. We present a summary of the combinatorial patterns observed within and between the 6 pathways examined.

A BRIEF OVERVIEW OF CANCER PATHWAYS

Figure 1 shows the genes in the combinatorial patterns and the simplified pathways in which they are involved. The cell cycle is regulated by several pairs of cyclins/cyclin dependent kinases. The decision to enter cell division or remain in a stationary phase is controlled by cyclin D/CDK4/6 and cyclin E/CDK2 pairs at the G_1 phase. CDK4 is inhibited by various proteins, such as TP15, TP16 (18, 19), and TP21. Both cyclin D/CDK4/6 and cyclin E/CDK2 complexes activate E2F, a master transcription factor at the G_1 phase (20). E2F is also inhibited by retinoblastoma (RB1) protein. The synthesis of cyclin D is regulated by several signal transduction pathways including Ras, Wnt, TGF-β, and IGF-AKT.

The Ras pathway is a MAP-kinase signaling pathway constituting families of growth factors, receptors, *e.g.*, EGFR and ERBB in lung (9–11), PDGFRA and KIT in gastrointestinal soft tissues (21), and KIT in hematopoietic tissues (22), G-proteins (KRAS, ref. 23; NRAS, ref. 24; and HRAS, ref. 25), receptor tyrosine kinases (*e.g.*, FLT3; refs. 5, 6), protein kinases (*e.g.*, BRAF;

TABLE 1. *Mutational patterns of single genes across 17 cancer tissues*

Gene	Acute lymphoblastic leukemia	Acute myeloid leukemia	Lung large cell carcinoma	Lung non small cell carcinoma	Lung adenocarcinoma	Lung squamous cell carcinoma	Lung small cell carcinoma	Pancreas
TP16	377 1123	56 679	13 50	86 676	85 346	66 277	6 277	196 616
	33.57%	8.25%	26.00%	12.72%	24.57%	23.83%	2.17%	31.82%
KRAS	80 1211	62 1382	38 168	311 1753	1001 4348	72 1286	4 329	2563 4405
	6.61%	4.49%	22.62%	17.74%	23.02%	5.60%	1.22%	58.18%
B-CATENIN	0 31	0 31	0 24	0 169	13 271	1 143	2 238	74 346
	0.00%	0.00%	0.00%	0.00%	4.80%	0.70%	0.84%	21.39%
NRAS	138 1454	336 2665	3 52	5 248	8 796	1 362	0 187	4 213
	9.49%	12.61%	5.77%	2.02%	1.01%	0.28%	0.00%	1.88%
TP53	14 31	10 31	4 17	4 7	29 56	6 18	43 72	12 18
	45.16%	32.26%	23.53%	57.14%	51.79%	33.33%	59.72%	66.67%
BRAF	2 62	2 255	0 31	3 179	13 510	6 229	0 98	5 198
	3.23%	0.78%	0.00%	1.68%	2.55%	2.62%	0.00%	2.53%
HRAS	1 448	0 925	2 53	1 235	1 511	3 263	0 166	0 189
	0.22%	0.00%	3.77%	0.43%	0.20%	1.14%	0.00%	0.00%
PIK3CA	1 31	0 99	1 22	3 249	2 44	0 111	3 68	4 66
	3.23%	0.00%	4.55%	1.20%	4.55%	0.00%	4.41%	6.06%
CDH1	0 31	0 31	0 11	0 6	1 44	0 12	0 68	0 20
	0.00%	0.00%	0.00%	0.00%	2.27%	0.00%	0.00%	0.00%
KIT	0 12	89 640	0 6	0 1	0 56	0 37	5 158	0 23
	0.00%	13.91%	0.00%	0.00%	0.00%	0.00%	3.16%	0.00%
PTEN	9 34	3 172	2 17	7 116	1 101	4 49	30 246	1 67
	26.47%	1.74%	11.76%	6.03%	0.99%	8.16%	12.20%	1.49%
RB1	1 31	3 102	0 18	3 255	3 99	2 33	53 116	0 32
	3.23%	2.94%	0.00%	1.18%	3.03%	6.06%	45.69%	0.00%
ABL1	22 66	148 558	1 6	0 1	0 13	0 7	0 4	0 1
	33.33%	26.52%	16.67%	0.00%	0.00%	0.00%	0.00%	0.00%
APC	0 73	0 67	2 13	0 6	2 145	4 68	1 144	24 152
	0.00%	0.00%	15.38%	0.00%	1.38%	5.88%	0.69%	15.79%
ATM	1 72	0 73	1 16	1 1	5 42	0 18	0 5	0 1
	1.39%	0.00%	6.25%	100.00%	11.90%	0.00%	0.00%	0.00%
EGFR	0 41	0 32	4 154	716 3783	1319 3679	35 1135	7 121	1 165
	0.00%	0.00%	2.60%	18.93%	35.85%	3.08%	5.79%	0.61%
FGFR3	0 8	0 1	1 6	0 1	1 19	0 18	0 9	0 1
	0.00%	0.00%	16.67%	0.00%	5.26%	0.00%	0.00%	0.00%
GATA1	0 12	61 137	0 0	0 0	0 0	0 0	0 0	0 0
	0.00%	44.53%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
JAK2	2 515	74 1116	0 6	0 263	0 13	0 7	0 45	0 1
	0.39%	6.63%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
SMAD4	0 31	2 123	0 15	0 6	4 61	1 19	1 80	135 603
	0.00%	1.63%	0.00%	0.00%	6.56%	5.26%	1.25%	22.39%
PTPN11	34 769	72 1803	0 0	0 1	1 124	0 0	0 5	0 1
	4.42%	3.99%	0.00%	0.00%	0.81%	0.00%	0.00%	0.00%
SMARCB1	5 73	1 61	0 0	0 34	0 6	0 0	0 22	1 2
	6.85%	1.64%	0.00%	0.00%	0.00%	0.00%	0.00%	50.00%
BRCA1	0 2	0 2	0 2	0 3	0 9	0 1	0 5	0 1
	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
CEBPA	0 23	199 1989	0 0	1 36	0 0	0 0	0 0	0 0
	0.00%	10.01%	0.00%	2.78%	0.00%	0.00%	0.00%	0.00%
FLT3	43 1176	3330 13436	1 8	0 3	0 16	0 8	0 4	0 1
	3.66%	24.78%	12.50%	0.00%	0.00%	0.00%	0.00%	0.00%
MEN1	0 0	0 0	0 0	0 0	0 0	0 0	1 88	31 144
	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	1.14%	21.53%
MSH6	1 7	2 12	0 0	0 1	0 6	0 0	0 5	0 19
	14.29%	16.67%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
NOTCH1	106 281	1 12	0 0	1 1	0 6	0 0	0 4	0 1
	37.72%	8.33%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%
NPM1	0 7	1169 3696	0 0	0 1	0 6	0 0	0 4	0 1
	0.00%	31.63%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
PDGFRA	0 11	4 131	0 8	1 3	0 59	0 38	0 61	0 23
	0.00%	3.05%	0.00%	33.33%	0.00%	0.00%	0.00%	0.00%
PTCH	0 0	0 0	0 0	0 0	0 0	0 6	0 0	0 0
	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
RET	0 8	0 0	0 6	0 1	1 29	0 11	2 25	0 18
	0.00%	0.00%	0.00%	0.00%	3.45%	0.00%	8.00%	0.00%
VHL	0 31	0 31	0 20	0 38	0 86	0 53	1 105	4 106
	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.95%	3.77%
WT1	3 38	16 193	0 1	0 0	0 27	0 10	0 0	0 0
	7.89%	8.29%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
CDKN2B	0 0	0 0	0 0	0 1	0 6	0 0	0 4	0 1
	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
CSF1R	1 11	9 284	0 6	0 1	1 14	1 8	0 4	0 1
	9.09%	3.17%	0.00%	0.00%	7.14%	12.50%	0.00%	0.00%
ERBB2	0 37	0 106	0 60	2 1085	44 1499	2 424	0 66	0 30
	0.00%	0.00%	0.00%	0.18%	2.94%	0.47%	0.00%	0.00%
FBXW7	0 0	0 0	0 0	0 1	0 56	0 0	0 4	0 13
	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
MET	0 8	0 0	0 6	2 10	2 140	0 7	6 46	0 17
	0.00%	0.00%	0.00%	20.00%	1.43%	0.00%	13.04%	0.00%
MPL	0 0	3 129	0 0	0 1	0 6	0 0	0 4	0 1
	0.00%	2.33%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
NRBP2	0 8	0 0	0 6	0 1	0 13	0 7	0 4	0 1
	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
SMO	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0
	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
LKB1	1 39	0 31	2 20	1 8	23 114	1 44	0 78	6 150
	2.56%	0.00%	10.00%	12.50%	20.18%	2.27%	0.00%	4.00%
TTN	0 8	0 0	3 6	1 1	1 13	2 8	3 4	0 1
	0.00%	0.00%	50.00%	100.00%	7.69%	25.00%	75.00%	0.00%

Top row of each entry: number of mutated and probed samples. Bottom row: mutational frequency.

TABLE 1. (*continued*)

Large intestine	Small intestine	Thyroid	Skin	Central nervous system	Soft tissue	Breast	Urinary tract	Endometrium
8 273	1 21	18 146	272 1191	262 1244	53 498	30 578	226 1371	12 233
2.93%	4.76%	12.33%	22.84%	21.06%	10.64%	5.19%	16.48%	5.15%
4207 13571	61 300	64 2474	31 1293	4 300	65 580	16 302	12 328	221 1570
31.00%	20.33%	2.59%	2.40%	1.33%	11.21%	5.30%	3.66%	14.08%
178 3162	13 61	29 174	83 596	40 645	84 567	0 340	5 226	184 867
5.63%	21.31%	16.67%	13.93%	6.20%	14.81%	0.00%	2.21%	21.22%
11 404	1 4	174 2628	606 3144	6 310	12 190	1 178	7 219	1 283
2.72%	25.00%	6.62%	19.27%	1.94%	6.32%	0.56%	3.20%	0.35%
20 71	0 1	10 12	16 47	32 68	8 19	32 70	14 20	8 10
28.17%	0.00%	83.33%	34.04%	47.06%	42.11%	45.71%	70.00%	80.00%
728 5288	1 41	1752 4936	1473 3603	15 385	9 279	4 143	0 196	5 402
13.77%	2.44%	35.49%	40.88%	3.90%	3.23%	2.80%	0.00%	1.24%
2 501	0 4	92 2460	89 1583	0 259	32 442	3 231	125 1071	4 293
0.40%	0.00%	3.74%	5.62%	0.00%	7.24%	1.30%	11.67%	1.37%
128 582	0 1	1 159	4 149	43 807	1 19	256 981	4 18	28 76
21.99%	0.00%	0.63%	2.68%	5.33%	5.26%	26.10%	22.22%	36.84%
1 49	0 1	1 39	0 47	0 59	20 79	110 541	2 49	3 82
2.04%	0.00%	2.56%	0.00%	0.00%	25.32%	20.33%	4.08%	3.66%
3 175	0 0	0 0	2 133	0 167	975 2736	1 85	0 9	0 21
1.71%	0.00%	0.00%	1.50%	0.00%	35.64%	1.18%	0.00%	0.00%
26 338	0 1	23 587	89 552	532 2738	7 155	30 558	12 141	483 1393
7.69%	0.00%	3.92%	16.12%	19.43%	4.52%	5.38%	8.51%	34.67%
1 78	0 1	0 12	4 69	23 249	7 103	7 65	18 54	4 16
1.28%	0.00%	0.00%	5.80%	9.24%	6.80%	10.77%	33.33%	25.00%
0 66	0 0	0 0	1 6	0 9	0 0	0 25	0 2	0 0
0.00%	0.00%	0.00%	16.67%	0.00%	0.00%	0.00%	0.00%	0.00%
1227 3596	31 211	17 206	3 133	6 392	12 168	18 467	0 43	1 48
34.12%	14.69%	8.25%	2.26%	1.53%	7.14%	3.85%	0.00%	2.08%
9 122	0 0	0 0	0 6	0 9	0 1	2 356	0 2	0 0
7.38%	0.00%	0.00%	0.00%	0.00%	0.00%	0.56%	0.00%	0.00%
4 806	0 49	2 89	2 106	38 834	0 317	1 259	0 21	0 19
0.50%	0.00%	2.25%	1.89%	4.56%	0.00%	0.39%	0.00%	0.00%
3 125	0 0	0 0	24 89	0 28	0 0	0 32	454 911	0 0
2.40%	0.00%	0.00%	26.97%	0.00%	0.00%	0.00%	49.84%	0.00%
0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0
0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
0 211	0 0	0 0	0 6	0 49	0 0	0 146	0 17	0 0
0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
87 842	8 47	15 81	1 51	1 62	0 20	6 130	1 27	0 10
10.33%	17.02%	18.52%	1.96%	1.61%	0.00%	4.62%	3.70%	0.00%
1 238	0 0	0 85	0 56	0 77	1 50	0 109	0 2	0 0
0.42%	0.00%	0.00%	0.00%	0.00%	2.00%	0.00%	0.00%	0.00%
0 23	0 0	0 1	0 9	72 574	107 248	0 145	0 2	1 1
0.00%	0.00%	0.00%	0.00%	12.54%	43.15%	0.00%	0.00%	100.00%
0 8	0 0	0 0	0 14	0 7	0 0	1 16	1 2	0 0
0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	6.25%	50.00%	0.00%
0 0	0 0	0 0	0 0	0 12	0 52	0 22	0 0	0 0
0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
0 73	0 0	0 0	0 14	0 34	0 28	0 32	0 2	0 0
0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
0 17	12 34	0 91	1 88	3 45	4 29	0 24	0 0	0 0
0.00%	35.29%	0.00%	1.14%	6.67%	13.79%	0.00%	0.00%	0.00%
48 158	0 0	0 1	1 6	2 30	0 0	0 9	5 90	14 135
30.38%	0.00%	0.00%	16.67%	6.67%	0.00%	0.00%	5.56%	10.37%
0 1	0 0	0 0	0 6	0 1	0 0	1 9	0 2	0 0
0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	11.11%	0.00%	0.00%
0 1	0 0	0 0	1 6	0 1	0 0	0 9	0 2	0 0
0.00%	0.00%	0.00%	16.67%	0.00%	0.00%	0.00%	0.00%	0.00%
0 228	0 0	0 0	1 46	7 333	281 1441	0 72	0 9	0 7
0.00%	0.00%	0.00%	2.17%	2.10%	19.50%	0.00%	0.00%	0.00%
1 11	0 0	0 0	122 438	23 240	0 15	2 52	0 53	0 0
9.09%	0.00%	0.00%	27.85%	9.58%	0.00%	3.85%	0.00%	0.00%
2 66	0 0	205 568	0 16	0 28	0 4	0 25	1 2	0 0
3.03%	0.00%	36.09%	0.00%	0.00%	0.00%	0.00%	50.00%	0.00%
10 238	0 1	1 44	0 72	2 97	23 108	1 187	0 42	0 14
4.20%	0.00%	2.27%	0.00%	2.06%	21.30%	0.53%	0.00%	0.00%
0 0	0 0	0 0	0 1	0 19	0 22	0 54	0 30	0 23
0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
0 1	0 0	0 0	0 6	0 1	0 0	0 9	0 2	0 0
0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
0 66	0 0	0 0	0 6	0 9	0 0	1 26	0 2	0 0
0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	3.85%	0.00%	0.00%
3 208	0 0	0 11	0 21	1 136	0 72	5 422	0 50	0 0
1.44%	0.00%	0.00%	0.00%	0.74%	0.00%	1.18%	0.00%	0.00%
44 533	0 0	0 0	1 6	0 26	0 0	1 17	0 2	0 0
8.26%	0.00%	0.00%	16.67%	0.00%	0.00%	5.88%	0.00%	0.00%
1 96	0 0	2 311	0 26	1 63	0 3	1 98	0 18	0 0
1.04%	0.00%	0.64%	0.00%	1.59%	0.00%	1.02%	0.00%	0.00%
0 1	0 0	0 0	0 6	0 1	0 0	0 9	0 2	0 0
0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
0 31	0 0	0 0	0 6	1 30	0 0	0 25	0 2	0 0
0.00%	0.00%	0.00%	0.00%	3.33%	0.00%	0.00%	0.00%	0.00%
5 8	0 0	0 0	17 156	1 39	0 0	0 48	0 0	0 0
62.50%	0.00%	0.00%	10.90%	2.56%	0.00%	0.00%	0.00%	0.00%
12 437	1 29	0 13	14 141	0 67	0 42	0 154	0 18	0 11
2.75%	3.45%	0.00%	9.93%	0.00%	0.00%	0.00%	0.00%	0.00%
9 31	0 0	0 0	4 6	0 9	0 1	4 80	2 2	0 0
29.03%	0.00%	0.00%	66.67%	0.00%	0.00%	5.00%	100.00%	0.00%

TABLE 2. *Mutational patterns of single genes in tumors and cell lines*

Gene	Acute lymphoblastic leukemia	Acute myeloid leukemia	Lung large cell carcinoma	Lung non small cell carcinoma	Lung adenocarcinoma	Lung squamous cell carcinoma	Lung small cell carcinoma
TP16	360 1092	45 648	7 39	84 670	67 302	59 266	6 209
	17 31	11 31	6 11	2 6	18 44	7 11	0 68
KRAS	78 1180	60 1351	33 156	310 1747	985 4302	70 1275	2 261
	2 31	2 31	5 12	1 6	16 46	2 11	2 68
B-CATENIN	0 0	0 0	0 13	0 163	11 227	1 132	1 170
	0 31	0 31	0 11	0 6	2 44	0 11	1 68
NRAS	130 1423	330 2634	2 41	4 241	7 752	1 351	0 119
	8 31	6 31	1 11	1 7	1 44	0 11	0 68
TP53	0 0	0 0	0 2	0 0	0 4	0 2	0 0
	14 31	10 31	4 15	4 7	29 52	6 16	43 72
BRAF	2 23	1 224	0 16	2 173	9 461	6 213	0 30
	0 39	1 31	0 15	1 6	4 49	0 16	0 68
HRAS	1 417	0 894	2 42	1 229	1 467	2 252	0 98
	0 31	0 31	0 11	0 6	0 44	1 11	0 68
PIK3CA	0 0	0 68	0 11	3 243	0 0	0 100	0 0
	1 31	0 31	1 11	0 6	2 44	0 11	3 68
CDH1	0 0	0 0	0 0	0 0	0 0	0 1	0 0
	0 31	0 31	0 11	0 6	1 44	0 11	0 68
KIT	0 4	89 640	0 2	0 0	0 47	0 32	5 154
	0 8	0 0	0 4	0 1	0 9	0 5	0 4
PTEN	1 3	2 141	1 6	6 110	1 57	3 38	22 178
	8 31	1 31	1 11	1 6	0 44	1 11	8 68
RB1	0 0	0 71	0 7	2 249	1 55	2 22	16 47
	1 31	3 31	0 11	1 6	2 44	0 11	37 69
ABL1	22 58	148 558	0 2	0 0	0 4	0 2	0 0
	0 8	0 0	1 4	0 1	0 9	0 5	0 4
APC	0 42	0 36	0 2	0 0	0 101	2 57	1 76
	0 31	0 31	2 11	0 6	2 44	2 11	0 68
ATM	1 64	0 73	0 12	0 0	2 33	0 13	0 0
	0 8	0 0	1 4	1 1	3 9	0 5	0 5
EGFR	0 2	0 1	4 139	716 3777	1317 3632	35 1119	7 53
	0 39	0 31	0 15	0 6	2 47	0 16	0 68
FGFR3	0 0	0 1	0 2	0 0	0 9	0 13	0 5
	0 8	0 0	1 4	0 1	1 10	0 5	0 4
GATA1	0 12	61 137	0 0	0 0	0 0	0 0	0 0
	0 0	0 0	0 0	0 0	0 0	0 0	0 0
JAK2	2 507	74 1116	0 2	0 262	0 4	0 2	0 41
	0 8	0 0	0 4	0 1	0 9	0 5	0 4
SMAD4	0 0	2 92	0 4	0 0	2 17	1 8	0 12
	0 31	0 31	0 11	0 6	2 44	0 11	1 68
PTPN11	34 769	72 1803	0 0	0 0	1 118	0 0	0 0
	0 0	0 0	0 0	0 1	0 6	0 0	0 5
SMARCB1	5 73	1 61	0 0	0 33	0 0	0 0	0 17
	0 0	0 0	0 0	0 1	0 6	0 0	0 5
BRCA1	0 0	0 0	0 0	0 0	0 0	0 0	0 0
	0 2	0 2	0 2	0 3	0 9	0 1	0 5
CEBPA	0 23	199 1989	0 0	1 36	0 0	0 0	0 0
	0 0	0 0	0 0	0 0	0 0	0 0	0 0
FLT3	42 1166	3330 13434	0 2	0 0	0 4	0 2	0 0
	1 10	0 2	1 6	0 3	0 12	0 6	0 4
MEN1	0 0	0 0	0 0	0 0	0 0	0 0	1 88
	0 0	0 0	0 0	0 0	0 0	0 0	0 0
MSH6	1 7	2 12	0 0	0 0	0 0	0 0	0 0
	0 0	0 0	0 0	0 1	0 6	0 0	0 5
NOTCH1	106 281	1 12	0 0	0 0	0 0	0 0	0 0
	0 0	0 0	0 0	1 1	0 6	0 0	0 4
NPM1	0 7	1169 3696	0 0	0 0	0 0	0 0	0 0
	0 0	0 0	0 0	0 1	0 6	0 0	0 4
PDGFRA	0 1	3 129	0 2	0 0	0 47	0 32	0 57
	0 10	1 2	0 6	1 3	0 12	0 6	0 4
PTCH	0 0	0 0	0 0	0 0	0 0	0 6	0 0
	0 0	0 0	0 0	0 0	0 0	0 0	0 0
RET	0 0	0 0	0 2	0 0	0 20	0 6	2 21
	0 8	0 0	0 4	0 1	1 9	0 5	0 4
VHL	0 0	0 0	0 9	0 32	0 42	0 42	1 37
	0 31	0 31	0 11	0 6	0 44	0 11	0 68
WT1	3 38	16 193	0 1	0 0	0 27	0 10	0 0
	0 0	0 0	0 0	0 0	0 0	0 0	0 0
CDKN2B	0 0	0 0	0 0	0 0	0 0	0 0	0 0
	0 0	0 0	0 0	0 1	0 6	0 0	0 4
CSF1R	1 3	9 284	0 2	0 0	0 4	0 2	0 0
	0 8	0 0	0 4	0 1	1 10	1 6	0 4
ERBB2	0 27	0 103	0 50	2 1080	44 1464	2 413	0 20
	0 10	0 3	0 10	0 5	0 35	0 11	0 46
FBXW7	0 0	0 0	0 0	0 0	0 50	0 0	0 0
	0 0	0 0	0 0	0 1	0 6	0 0	0 4
MET	0 0	0 0	0 2	1 9	2 131	0 2	6 42
	0 8	0 0	0 4	1 1	0 9	0 5	0 4
MPL	0 0	3 129	0 0	0 0	0 0	0 0	0 0
	0 0	0 0	0 0	0 1	0 6	0 0	0 4
NRBP2	0 0	0 0	0 2	0 0	0 4	0 2	0 0
	0 8	0 0	0 4	0 1	0 9	0 5	0 4
SMO	0 0	0 0	0 0	0 0	0 0	0 0	0 0
	0 0	0 0	0 0	0 0	0 0	0 0	0 0
LKB1	0 0	0 0	0 5	0 2	17 66	0 28	0 10
	1 39	0 31	2 15	1 6	6 48	1 16	0 68
TTN	0 0	0 0	1 2	0 0	0 4	0 2	0 0
	0 8	0 0	2 4	1 1	1 9	2 6	3 4

Top row of each entry: mutated and probed samples in primary tumors. Bottom row: corresponding numbers in cancer cell lines.

TABLE 2. (*continued*)

Pancreas	Large intestine	Small intestine	Thyroid	Skin	Central nervous system	Soft tissue	Breast	Urinary tract	Endometrium
187 599	5 233	1 20	16 134	244 1144	238 1185	47 479	22 533	216 1353	12 223
9 17	3 40	0 1	2 12	28 47	24 59	6 19	8 45	10 18	0 10
2547 4388	4128 13410	61 299	62 2462	31 1246	4 241	65 561	15 257	11 310	219 1560
16 17	79 161	0 1	2 12	0 47	0 59	0 19	1 45	1 18	2 10
74 329	175 3122	12 60	29 162	82 549	40 586	84 548	0 295	5 208	183 857
0 17	3 40	1 1	0 12	1 47	0 59	0 19	0 45	0 18	1 10
4 196	11 364	1 3	174 2616	596 3097	5 251	9 171	1 133	4 200	1 273
0 17	0 40	0 1	0 12	10 47	1 59	3 19	0 45	3 19	0 10
0 0	0 0	0 0	0 0	0 0	0 7	0 0	0 0	0 0	0 0
12 18	20 71	0 1	10 12	16 47	32 61	8 19	32 70	14 20	8 10
5 181	714 5127	1 40	1747 4924	1441 3553	12 325	5 260	1 82	0 178	5 392
0 17	14 161	0 1	5 12	32 50	3 60	4 19	3 61	0 18	0 10
0 172	2 461	0 3	92 2448	88 1536	0 200	32 423	2 186	123 1053	3 283
0 17	0 40	0 1	0 12	1 47	0 59	0 19	1 45	2 18	1 10
4 49	116 542	0 0	0 147	4 102	42 748	0 0	245 936	0 0	24 66
0 17	12 40	0 1	1 12	0 47	1 59	1 19	11 45	4 18	4 10
0 3	1 9	0 0	1 27	0 0	0 0	20 60	108 496	2 31	3 72
0 17	0 40	0 1	0 12	0 47	0 59	0 19	2 45	0 18	0 10
0 22	0 53	0 0	0 0	2 127	0 165	975 2736	1 60	0 7	0 21
0 1	3 122	0 0	0 0	0 6	0 2	0 0	0 25	0 2	0 0
1 50	25 298	0 0	22 575	81 505	507 2677	6 136	22 513	8 123	480 1383
0 17	1 40	0 1	1 12	8 47	25 61	1 19	8 45	4 18	3 10
0 15	0 38	0 0	0 0	3 22	13 190	7 84	2 20	11 36	1 6
0 17	1 40	0 1	0 12	1 47	10 59	0 19	5 45	7 18	3 10
0 0	0 35	0 0	0 0	0 0	0 7	0 0	0 0	0 0	0 0
0 1	0 31	0 0	0 0	1 6	0 2	0 0	0 25	0 2	0 0
24 135	1200 3556	31 210	17 194	2 86	5 333	11 149	17 422	0 25	1 38
0 17	27 40	0 1	0 12	1 47	1 59	1 19	1 45	0 18	0 10
0 0	0 0	0 0	0 0	0 0	0 7	0 1	1 331	0 0	0 0
0 1	9 122	0 0	0 0	0 6	0 2	0 0	1 25	0 2	0 0
1 148	3 736	0 48	2 77	1 59	38 774	0 298	1 198	0 3	0 9
0 17	1 70	0 1	0 12	1 47	0 60	0 19	0 61	0 18	0 10
0 0	2 94	0 0	0 0	24 83	0 26	0 0	0 7	453 909	0 0
0 1	1 31	0 0	0 0	0 6	0 2	0 0	0 25	1 2	0 0
0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0
0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0
0 0	0 180	0 0	0 0	0 0	0 47	0 0	0 121	0 15	0 0
0 1	0 31	0 0	0 0	0 6	0 2	0 0	0 25	0 2	0 0
129 586	81 802	8 46	15 69	0 4	0 3	0 1	5 85	1 9	0 0
6 17	6 40	0 1	0 12	1 47	1 59	0 19	1 45	0 18	0 10
0 0	1 237	0 0	0 85	0 50	0 76	1 50	0 100	0 0	0 0
0 1	0 1	0 0	0 0	0 6	0 1	0 0	0 9	0 2	0 0
1 1	0 22	0 0	0 1	0 3	72 573	107 248	0 136	0 0	1 1
0 1	0 1	0 0	0 0	0 6	0 1	0 0	0 9	0 2	0 0
0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0
0 1	0 8	0 0	0 0	0 14	0 7	0 0	1 16	1 2	0 0
0 0	0 0	0 0	0 0	0 0	0 12	0 52	0 22	0 0	0 0
0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0
0 0	0 35	0 0	0 0	0 0	0 26	0 28	0 0	0 0	0 0
0 1	0 38	0 0	0 0	0 14	0 8	0 0	0 32	0 2	0 0
31 144	0 17	12 34	0 91	1 88	3 45	4 29	0 24	0 0	0 0
0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0
0 18	48 157	0 0	0 1	0 0	1 28	0 0	0 0	5 88	14 135
0 1	0 1	0 0	0 0	1 6	1 2	0 0	0 9	0 2	0 0
0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0
0 1	0 1	0 0	0 0	0 6	0 1	0 0	1 9	0 2	0 0
0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0
0 1	0 1	0 0	0 0	1 6	0 1	0 0	0 9	0 2	0 0
0 22	0 190	0 0	0 0	0 32	6 325	281 1441	0 40	0 7	0 7
0 1	0 38	0 0	0 0	1 14	1 8	0 0	0 32	0 2	0 0
0 0	1 11	0 0	0 0	122 438	23 240	0 15	2 52	0 53	0 0
0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0
0 17	0 35	0 0	205 568	0 10	0 26	0 4	0 0	0 0	0 0
0 1	2 31	0 0	0 0	0 6	0 2	0 0	0 25	1 2	0 0
4 89	10 198	0 0	1 32	0 25	2 38	23 89	0 142	0 24	0 4
0 17	0 40	0 1	0 12	0 47	0 59	0 19	1 45	0 18	0 10
0 0	0 0	0 0	0 0	0 1	0 19	0 22	0 54	0 30	0 23
0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0
0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0
0 1	0 1	0 0	0 0	0 6	0 1	0 0	0 9	0 2	0 0
0 0	0 35	0 0	0 0	0 0	0 7	0 0	0 0	0 0	0 0
0 1	0 31	0 0	0 0	0 6	0 2	0 0	1 26	0 2	0 0
0 25	3 170	0 0	0 11	0 7	1 120	0 70	4 390	0 39	0 0
0 5	0 38	0 0	0 0	0 14	0 16	0 2	1 32	0 11	0 0
0 12	44 532	0 0	0 0	0 0	0 25	0 0	1 8	0 0	0 0
0 1	0 1	0 0	0 0	1 6	0 1	0 0	0 9	0 2	0 0
0 16	1 65	0 0	2 311	0 20	1 61	0 3	0 73	0 16	0 0
0 1	0 31	0 0	0 0	0 6	0 2	0 0	1 25	0 2	0 0
0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0
0 1	0 1	0 0	0 0	0 6	0 1	0 0	0 9	0 2	0 0
0 0	0 0	0 0	0 0	0 0	0 28	0 0	0 0	0 0	0 0
0 1	0 31	0 0	0 0	0 6	1 2	0 0	0 25	0 2	0 0
0 0	5 8	0 0	0 0	17 156	1 39	0 0	0 48	0 0	0 0
0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0
6 133	11 367	1 28	0 1	12 93	0 7	0 23	0 93	0 0	0 1
0 17	1 70	0 1	0 12	2 48	0 60	0 19	0 61	0 18	0 10
0 0	0 0	0 0	0 0	0 0	0 7	0 1	0 55	0 0	0 0
0 1	9 31	0 0	0 0	4 6	0 2	0 0	4 25	2 2	0 0

TABLE 3. *Mutational frequencies demonstrating significant differences between cell line and tumor data*

Gene	Tissue	Non-cell line frequency	Cell line frequency	*P* value
TP16	Central nervous system	238 1185	24 59	4.54E-006
KRAS	Large intestine	4125 13,410	79 161	3.85E-009
BRAF	Large intestine	682 5127	14 161	3.46E-004
BRAF	Skin	1374 3553	32 50	2.34E-006
PTEN	Central nervous system	478 2677	24 61	9.12E-007
RB1	Central nervous system	7 190	10 59	6.98E-006

refs. 26, 27), protein phosphatases (*e.g.*, PTPN11; ref. 28), and transcription factors (*e.g.*, CEBPA; ref. 29). The Wnt pathway is important in development and tissue regeneration and is often altered in gastrointestinal tumors (30). Two genes in the Wnt pathway undergo frequent mutations: β-catenin and APC. APC, GSK-3β, and other proteins bind and phosphorylate β-catenin, targeting its degradation in the absence of the WNT ligand (30).

The IGF-AKT pathway transduces the survival signal in response to growth factor stimulation (31). The pathway is activated when PIP_3 is phosphorylated by PI3

TABLE 4. *Somatic mutation type statistics*

Gene	Function	Substitution	Insertion	Deletion	Frameshift	Other	Total
TP16	Tumor suppressor	584	7	964	128	1146	2,829
KRAS	Oncogene	10,032	6	0	2	0	10,040
B-CATENIN	Oncogene	1,382	1	182	3	16	1,584
NRAS	Oncogene	1,631	1	1	1	0	1,634
TP53	Tumor suppressor	351	3	6	34	25	419
BRAF	Oncogene	4,137	3	1	0	2	4,143
HRAS	Oncogene	520	0	0	1	0	521
PIK3CA	Oncogene	574	2	3	14	4	597
CDH1	Tumor suppressor	91	0	14	62	34	201
KIT	Oncogene	834	173	406	1	20	1,434
PTEN	Tumor suppressor	682	7	48	429	263	1,429
RB1	Tumor suppressor	112	2	3	53	79	249
ABL1	Oncogene	173	0	0	0	0	173
APC	Tumor suppressor	520	0	0	868	65	1,453
ATM	Tumor suppressor	127	1	4	29	5	166
EGFR	Oncogene	1,304	46	728	3	96	2,177
FGFR3	Oncogene	506	2	0	3	0	511
GATA1	Tumor suppressor	27	1	1	75	1	105
JAK2	Oncogene	3,823	0	4	0	0	3,827
SMAD4	Tumor suppressor	125	0	3	26	133	287
PTPN11	Oncogene	284	0	0	0	0	284
SMARCB1	Tumor suppressor	60	0	69	40	21	190
BRCA1	Oncogene	189	0	0	1	0	190
CEBPA	Tumor suppressor	30	23	11	151	1	216
FLT3	Oncogene	227	2915	16	0	23	3,181
MEN1	Tumor suppressor	80	1	11	70	16	178
MSH6	Tumor suppressor	16	0	0	76	0	92
NOTCH1	Oncogene	88	3	6	12	0	109
NPM1	Tumor suppressor	1	0	0	876	241	1,118
PDGFRA	Oncogene	246	1	46	1	1	295
PTCH	Tumor suppressor	106	1	3	43	11	164
RET	Oncogene	220	0	7	0	0	227
VHL	Tumor suppressor	402	5	26	373	72	878
WT1	Both	33	0	19	38	19	109
CDKN2B	Tumor suppressor	7	0	11	1	6	25
CSF1R	Oncogene	36	0	0	0	0	36
ERBB2	Oncogene	44	32	0	0	0	76
FBXW7	Tumor suppressor	50	0	1	4	0	55
MET	Oncogene	63	0	1	0	3	67
MPL	Oncogene	20	0	0	0	0	20
NRBP2	Oncogene	34	0	0	0	0	34
SMO	Oncogene	25	1	0	0	0	26
LKB1	Tumor suppressor	58	0	0	20	9	87
TTN	Unknown	51	0	0	0	0	51

kinase PIK3CA and inhibited when it is dephosphorylated by PI3 phosphatase PTEN (32). In addition, PI3 kinase is also regulated by the Ras proteins (33, 34). The TGF-β signaling pathway imposes growth inhibition under stressful conditions. SMAD4 encodes a protein in the TGF-β pathway (35, 36). It activates the synthesis of TP15, which inhibits CDK4 (37).

The TP53 pathway responds to a variety of stress signals and controls major stress response processes including apoptosis, cell cycle arrest, and senescence (38, 39). To coordinate the stress response TP53 transcriptionally regulates genes in other pathways such as TP21 in the cell cycle control (40) and PTEN in the IGF-AKT pathway (41).

CELL CYCLE CONTROL, STRESS RESPONSES, IGF-AKT, AND TGF-β PATHWAYS

Mutations of genes in cell cycle control and stress response pathways are often found to co-occur in a sample. TP16 and TP53 have a significant co-occurring pattern in 11 different tissue types (see Table 5). For instance, in 44 lung adenocarcinoma samples where TP16 and TP53 are cosequenced, TP16 and TP53 are mutated in 18 and 29 samples, respectively, and both genes are mutated together in 13 samples (LR 1.1; $P \leq 8.25 \times 10^{-4}$). RB1 and TP53 have a significant co-occurring pattern in lung small cell carcinoma (LR 1.11; $P \leq 8.51 \times 10^{-6}$), central nervous system (LR 1.66; $P \leq 4.61 \times 10^{-3}$), and urinary tract (LR 1.29; $P \leq 6.14 \times 10^{-3}$) tumors. Comutations between stress response and IGF-AKT pathways and between cell cycle control and IGF-AKT pathways are also often observed. PTEN and TP53 have a co-occurring mutational pattern in 6 different tissue types. PIK3CA and TP53 have a co-occurring mutational pattern in large intestine (LR 0.67; $P \leq 3.45 \times 10^{-2}$), breast (LR 0.82; $P \leq 1.43 \times 10^{-2}$), and endometrium (LR 0.94; $P \leq 4.88 \times 10^{-2}$). TP16 and PTEN have a co-occurring mutational pattern in acute lymphoblastic leukemia (ALL) (LR 1.6; $P \leq 8.39 \times 10^{-3}$), skin (LR 1.05, $P \leq 7.62 \times 10^{-4}$), and central nervous system (LR 1.17; $P \leq 1.06 \times 10^{-3}$). Furthermore, three genes belonging to each pathway—TP16, TP53, PTEN—also demonstrate a significant (yet weaker) co-occurring pattern in ALL (LR 0.98; $P \leq 1.13 \times 10^{-2}$), central nervous system (LR 0.85; $P \leq 1.38 \times 10^{-2}$), and urinary tract (LR 1; $P \leq 4.87 \times 10^{-2}$) tumors.

Genes in the TGF-β pathway are also comutated in the same sample with genes in the cell cycle control and stress response pathways. SMAD4 and TP53 have a co-occurring mutational pattern in the pancreas (LR 0.94; $P \leq 3.04 \times 10^{-2}$), large intestine (LR 1.33, $P \leq 3.45 \times 10^{-2}$), and upper aerodigestive tract (LR 1.15, $P \leq 3.23 \times 10^{-2}$), and TP16 and SMAD4 have a co-occurring mutational pattern in the pancreas (LR 1.15, $P \leq 6.11 \times 10^{-5}$). The co-occurring mutations of the three genes are statistically insignificant due to the insufficient number of samples probing all the three genes. No significant co-occurrence or mutually exclusive pattern was found between the IGF-AKT (PTEN, PIK3CA) and TGF-β (SMAD4) genes, since SMAD4 mutations occur mainly in the pancreas and large intestine, where PTEN and PIK3CA mutations are rare.

RAS, CELL CYCLE CONTROL, STRESS RESPONSE, IGF-AKT, AND TGF-β PATHWAYS

Genes in the Ras pathway are commonly mutated in many tissue types and hence intersect with mutations in cell cycle control, stress response, IGF-AKT, and TGF-β pathways. TP16 and KRAS have a co-occurring pattern in 5 different tissue types. TP16 and NRAS have a co-occurring pattern in ALL (LR 1.14, $P \leq 2.08 \times 10^{-2}$) and skin (LR 0.87, $P \leq 2.10 \times 10^{-3}$). KRAS and TP53 have a co-occurring pattern in lung adenocarcinoma (LR 0.85, $P \leq 4.15 \times 10^{-3}$), pancreas (LR 0.97, $P \leq 4.44 \times 10^{-4}$), and large intestine (LR 1.06, $P \leq 3.35 \times 10^{-3}$). NRAS and TP53 have a co-occurring pattern in both ALL (LR 1.11, $P \leq 3.37 \times 10^{-2}$) and AML (LR 2.07, $P \leq 3.37 \times 10^{-2}$). BRAF and TP53 have a co-occurring pattern in large intestine (LR 1.75, $P \leq 3.59 \times 10^{-2}$), thyroid (LR 0.96, $P \leq 2.73 \times 10^{-2}$) and skin (LR 0.73, $P \leq 6.35 \times 10^{-3}$). Other significant co-occurring pairs or triplets include KRAS and SMAD4 (pancreas); KRAS and PTEN (endometrium); NRAS and PTEN (ALL); BRAF and TP16 (skin); BRAF and PTEN (skin); KRAS and PIK3CA (large intestine); TP16, KRAS and SMAD4 (pancreas); TP16, KRAS, and TP53 (pancreas); KRAS, SMAD4, and TP53 (pancreas); BRAF, TP16, and PTEN (skin); TP16, NRAS, and PTEN (ALL); and TP16, NRAS, and TP53 (ALL).

WNT, RAS, STRESS RESPONSE, IGF-AKT, AND TGF-β PATHWAYS

The two genes in the Wnt pathway—APC and β-catenin—are frequently mutated in gastrointestinal tumors (30). Co-occurring patterns of the Wnt genes with genes in Ras, stress response, IGF-AKT, and TGF-β pathways are observed. In large intestine, APC and KRAS demonstrate a strong co-occurring pattern (LR 1.23, $P \leq 3.41 \times 10^{-7}$). APC-BRAF (LR 0.95, $P \leq 2.27 \times 10^{-2}$), and β-catenin-KRAS (LR 1.13, $P \leq 5.78 \times 10^{-3}$) pairs also have co-occurring patterns. Large intestine samples also carry co-occurring patterns in the following pairs and triplets: APC and TP53; APC and SMAD4; APC and PIK3CA; APC, SMAD4, and TP53; APC, KRAS, and PIK3CA; and APC, KRAS, and TP53. No significant patterns between Wnt genes and cell cycle control genes were detected. This is mainly because the cell cycle genes (TP16, RB1) have rarely mutated or been cosequenced with APC or β-catenin in gastrointestinal tissues.

INTRAPATHWAY PATTERNS

Key genes of the Ras pathway have mutually exclusive patterns in many tissues: BRAF and NRAS (thyroid and

TABLE 5. *Combinatorial mutational patterns*

Gene						Gene mutated					
1	2	3	Tissue	Type	Cosequenced	1	2	3	All	Ratio	*P* value
PTEN	TP53	—	B cell lymphoma	Co-occurred	13	5	5	—	4	2.08	2.81E-002
KIT	KRAS	—	T cell lymphoma	Co-occurred	114	19	9	—	4	2.667	3.65E-002
B-CATENIN	KIT	—	T cell lymphoma acute lymphoblastic	Co-occurred	114	25	19	—	7	1.68	1.09E-002
TP16	TP53	—	Acute lymphoblastic leukemia	Co-occurred	31	17	14	—	9	1.172	3.44E-003
TP16	PTEN	—	Acute lymphoblastic leukemia	Co-occurred	31	17	8	—	7	1.596	8.39E-003
PTEN	TP53	—	Acute lymphoblastic leukemia	Co-occurred	31	8	14	—	7	1.938	8.39E-003
TP16	PTEN	TP53	Acute lymphoblastic leukemia	Co-occurred	31	17	8	14	6	0.98	1.13E-002
KRAS	NRAS	—	Acute lymphoblastic leukemia	Co-occurred	856	80	109	—	5	0.491	2.46E-002
NRAS	PTPN11	—	Acute lymphoblastic leukemia	Mutually exclusive	403	42	29	—	0	0	3.63E-002
NRAS	PTEN	—	Acute lymphoblastic leukemia	Co-occurred	31	8	8	—	4	1.938	3.37E-002
TP16	NRAS	—	Acute lymphoblastic leukemia	Co-occurred	31	17	8	—	5	1.14	2.08E-002
NRAS	TP53	—	Acute lymphoblastic leukemia	Co-occurred	31	8	14	—	4	1.107	3.37E-002
TP16	NRAS	PTEN	Acute lymphoblastic leukemia	Co-occurred	31	17	8	8	4	1.143	3.09E-002
TP16	NRAS	TP53	Acute lymphoblastic leukemia	Co-occurred	31	17	8	14	4	1.511	3.09E-002
KRAS	NRAS	—	Acute myeloid leukemia	Co-occurred	578	58	159	—	10	0.627	4.30E-003
CEBPA	FLT3	—	Acute myeloid leukemia	Mutually exclusive	647	29	186	—	3	0.36	1.54E-002
FLT3	NPM1	—	Acute myeloid leukemia	Co-occurred	110	35	50	—	18	1.131	2.74E-004
NRAS	TP53	—	Acute myeloid leukemia	Co-occurred	31	6	10	—	4	2.067	3.37E-002
FLT3	KIT	—	Acute myeloid leukae	Mutually exclusive	95	8	30	—	0	0	4.15E-002
JAK2	MPL	—	Hematopoietic and lymphoid others	Co-occurred	830	550	13	—	6	0.697	1.68E-002
TP16	KRAS	—	Lung large cell carcinoma	Co-occurred	11	6	5	—	3	1.1	4.98E-002
TP16	TP53	—	Lung adenocarcinoma	Co-occurred	44	18	29	—	13	1.096	8.25E-004
BRAF	KRAS	—	Lung adenocarcinoma	Mutually exclusive	72	6	40	—	0	0	5.80E-003
KRAS	NRAS	—	Lung adenocarcinoma	Mutually exclusive	194	127	7	—	0	0	4.73E-004
EGFR	KRAS	—	Lung adenocarcinoma	Mutually exclusive	641	279	122	—	2	0.038	1.05E-031
TP16	KRAS	—	Lung adenocarcinoma	Co-occurred	64	19	23	—	5	0.732	2.29E-002
KRAS	TP53	—	Lung adenocarcinoma	Co-occurred	44	16	29	—	9	0.853	4.15E-003
LKB1	TP53	—	Lung adenocarcinoma	Mutually exclusive	51	6	29	—	1	0.293	4.66E-002
TP16	TP53	—	Lung squamous cell carcinoma	Co-occurred	11	7	6	—	3	0.786	4.98E-002
PTEN	TP53	—	Lung small cell carcinoma	Co-occurred	68	8	43	—	5	0.988	2.30E-002
PTEN	RB1	—	Lung small cell carcinoma	Co-occurred	68	8	37	—	6	1.378	1.52E-002
RB1	TP53	—	Lung small cell carcinoma	Co-occurred	68	37	43	—	26	1.111	8.51E-006
PTEN	RB1	TP53	Lung small cell carcinoma	Co-occurred	68	8	37	43	4	0.949	3.45E-002
TP16	SMAD4	—	Pancreas	Co-occurred	89	50	34	—	22	1.152	6.11E-005
TP16	TP53	—	Pancreas	Co-occurred	17	9	12	—	7	1.102	5.85E-003
SMAD4	TP53	—	Pancreas	Co-occurred	17	6	12	—	4	0.944	3.04E-002
TP16	KRAS	—	Pancreas	Co-occurred	157	72	103	—	55	1.164	2.58E-009
KRAS	SMAD4	—	Pancreas	Co-occurred	157	87	37	—	24	1.171	5.41E-005
KRAS	TP53	—	Pancreas	Co-occurred	17	16	12	—	11	0.974	4.44E-004
TP16	KRAS	SMAD4	Pancreas	Co-occurred	89	50	64	34	15	0.852	5.57E-004
TP16	KRAS	TP53	Pancreas	Co-occurred	17	9	16	12	6	0.935	7.47E-003
KRAS	SMAD4	TP53	Pancreas	Co-occurred	17	16	6	12	3	0.818	4.80E-002
B-CATENIN	KRAS	—	Pancreas	Mutually exclusive	78	23	33	—	2	0.206	6.18E-005
BRAF	KRAS	—	Biliary tract	Mutually exclusive	50	15	33	—	3	0.303	1.56E-005
TP16	KRAS	—	Biliary tract	Co-occurred	40	22	5	—	4	1.455	3.45E-002
SMAD4	TP53	—	Large intestine	Co-occurred	40	6	20	—	4	1.333	3.45E-002
PIK3CA	TP53	—	Large intestine	Co-occurred	40	12	20	—	4	0.667	3.45E-002
BRAF	KRAS	—	Large intestine	Mutually exclusive	946	122	345	—	9	0.202	3.46E-015
APC	B-CATENIN	—	Large intestine	Mutually exclusive	485	230	10	—	1	0.211	1.53E-002
TP16	KRAS	—	Large intestine	Co-occurred	74	5	44	—	4	1.345	3.59E-002
KRAS	TP53	—	Large intestine	Co-occurred	70	33	20	—	10	1.061	3.35E-003
BRAF	TP53	—	Large intestine	Co-occurred	70	8	20	—	4	1.75	3.59E-002
KRAS	PIK3CA	—	Large intestine	Co-occurred	70	35	14	—	10	1.429	3.35E-003
APC	TP53	—	Large intestine	Co-occurred	40	27	20	—	15	1.111	3.12E-004
APC	SMAD4	—	Large intestine	Co-occurred	40	27	6	—	6	1.481	1.40E-002
APC	PIK3CA	—	Large intestine	Co-occurred	40	27	12	—	7	0.864	9.13E-003
APC	SMAD4	TP53	Large intestine	Co-occurred	40	27	6	20	4	1	3.24E-002
APC	KRAS	—	Large intestine	Co-occurred	389	121	118	—	45	1.226	3.41E-007
B-CATENIN	KRAS	—	Large intestine	Co-occurred	303	26	93	—	9	1.128	5.78E-003
APC	BRAF	—	Large intestine	Co-occurred	60	35	9	—	5	0.952	2.27E-002
APC	KRAS	PIK3CA	Large intestine	Co-occurred	40	27	22	12	5	0.859	1.99E-002
APC	KRAS	TP53	Large intestine	Co-occurred	40	27	22	20	9	1.012	3.26E-003
TP16	TP53	—	Esophagus	Co-occurred	23	8	19	—	6	0.908	1.19E-002
BRAF	NRAS	—	Thyroid	Mutually exclusive	280	184	6	—	0	0	1.46E-003
HRAS	NRAS	—	Thyroid	Mutually exclusive	1435	73	136	—	1	0.145	5.20E-003
BRAF	TP53	—	Thyroid	Co-occurred	12	5	10	—	4	0.96	2.73E-002
TP16	TP53	—	Skin	Co-occurred	47	28	16	—	6	0.629	1.44E-002
TP16	PTEN	—	Skin	Co-occurred	129	83	22	—	15	1.06	7.62E-004
BRAF	NRAS	—	Skin	Mutually exclusive	906	452	141	—	10	0.142	2.38E-032
BRAF	HRAS	—	Skin	Mutually exclusive	337	133	8	—	0	0	1.71E-002

(*continues*)

TABLE 5. (*continued*)

Gene						Gene mutated					
1	2	3	Tissue	Type	Cosequenced	1	2	3	All	Ratio	*P* value
BRAF	KRAS	—	Skin	Mutually exclusive	265	140	4	—	0	0	4.83E-002
KRAS	NRAS	—	Skin	Co-occurred	604	20	182	—	4	0.664	3.74E-002
BRAF	TP53	—	Skin	Co-occurred	47	32	16	—	8	0.734	6.35E-003
NRAS	PTEN	—	Skin	Mutually exclusive	160	34	35	—	2	0.269	6.44E-003
BRAF	TP16	—	Skin	Co-occurred	110	66	51	—	31	1.013	3.86E-006
BRAF	PTEN	—	Skin	Co-occurred	87	48	10	—	10	1.812	3.56E-003
TP16	NRAS	—	Skin	Co-occurred	175	83	29	—	12	0.872	2.10E-003
NRAS	TP53	—	Skin	Co-occurred	47	10	16	—	4	1.175	3.50E-002
BRAF	TP16	PTEN	Skin	Co-occurred	83	47	51	10	7	1	9.85E-003
TP16	TP53	—	Central nervous system	Co-occurred	59	24	32	—	13	0.999	1.04E-003
TP16	PTEN	—	Central nervous system	Co-occurred	137	42	39	—	14	1.171	1.06E-003
PTEN	TP53	—	Central nervous system	Co-occurred	67	25	32	—	16	1.34	3.80E-004
TP16	RB1	—	Central nervous system	Mutually exclusive	82	31	13	—	1	0.203	1.21E-002
PTEN	RB1	—	Central nervous system	Co-occurred	59	25	10	—	6	1.416	1.49E-002
RB1	TP53	—	Central nervous system	Co-occurred	59	10	32	—	9	1.659	4.61E-003
TP16	PTEN	TP53	Central nervous system	Co-occurred	59	24	25	32	6	0.852	1.38E-002
PTEN	RB1	TP53	Central nervous system	Co-occurred	59	25	10	32	5	0.926	2.14E-002
HRAS	KRAS	—	Soft tissue	Co-occurred	206	20	29	—	7	2.486	1.13E-002
KIT	PDGFRA	—	Soft tissue	Mutually exclusive	778	270	126	—	1	0.023	8.22E-025
TP16	TP53	—	Breast	Co-occurred	45	8	30	—	5	0.938	2.21E-002
PTEN	TP53	—	Breast	Co-occurred	45	8	30	—	6	1.125	1.43E-002
PIK3CA	TP53	—	Breast	Co-occurred	45	11	30	—	6	0.818	1.43E-002
TP16	TP53	—	Upper aerodigestive tract	Co-occurred	23	11	20	—	11	1.15	9.15E-004
SMAD4	TP53	—	Upper aerodigestive tract	Co-occurred	23	4	20	—	4	1.15	3.23E-002
TP16	TP53	—	Urinary tract	Co-occurred	18	10	14	—	7	0.9	6.14E-003
PTEN	TP53	—	Urinary tract	Co-occurred	18	4	14	—	4	1.286	3.08E-002
TP16	RB1	—	Urinary tract	Mutually exclusive	22	11	8	—	1	0.25	1.19E-002
RB1	TP53	—	Urinary tract	Co-occurred	18	7	14	—	7	1.286	6.14E-003
TP16	PTEN	TP53	Urinary tract	Co-occurred	18	10	4	14	3	1	4.87E-002
HRAS	KRAS	—	Prostate	Mutually exclusive	117	26	36	—	0	0	1.56E-005
KRAS	NRAS	—	Prostate	Mutually exclusive	109	35	8	—	0	0	3.97E-002
PIK3CA	TP53	—	Endometrium	Co-occurred	10	4	8	—	3	0.938	4.88E-002
KRAS	PTEN	—	Endometrium	Co-occurred	59	12	24	—	7	1.434	9.97E-003
TP16	TP53	—	Kidney	Co-occurred	22	13	7	—	4	0.967	3.20E-002
B-CATENIN	WT1	—	Kidney	Co-occurred	70	12	11	—	5	2.652	2.30E-002

skin), BRAF and KRAS (lung adenocarcinoma, biliary tract, large intestine, and skin), BRAF and HRAS (skin), HRAS and NRAS (thyroid), HRAS and KRAS (prostate), and KRAS and NRAS (lung adenocarcinoma and prostate). These key genes also form tissue-specific mutually exclusive patterns with selected receptors, kinases, phosphatases, or transcription factors in the Ras pathway. Examples include the pairs of KRAS and EGFR in lung adenocarcinoma (LR 0.04, $P \leq 1.05 \times 10^{-31}$), NRAS and PTPN11 in ALL (LR 0, $P \leq 3.63 \times 10^{-2}$), KIT and PDGFRA in gastrointestinal soft tissues (LR 0.02, $P \leq 8.22 \times 10^{-25}$) (21), FLT3 and CEBPA in AML (LR 0.36, $P \leq 1.54 \times 10^{-2}$) (42). Curiously, these pairs include both genes that have clear upstream-downstream relations (KRAS and EGFR, RAS genes and BRAF, and FLT3 and CEBPA) and genes that serve complementary functions in the Ras pathway (different RAS genes, KIT, and PDGFRA). The alteration of any of those genes seems sufficient to promote abnormal cell growth.

TABLE 6. *Number of mutational patterns connecting pathways*

	Cell cycle	Stress response	Ras	IGF-AKT	Wnt	TGFB
Co-occurrence						
Cell cycle	0	21	13	12	0	2
Stress response	21	0	13	14	3	5
Ras	13	13	6	7	6	3
IGF-AKT	12	14	7	0	2	0
Wnt	0	3	6	2	0	2
TGFB	2	5	3	0	2	0
Mutual exclusion						
Cell cycle	2	0	0	0	0	0
Stress response	0	0	0	0	0	0
Ras	0	0	16	1	1	0
IGF-AKT	0	0	1	0	0	0
Wnt	0	0	1	0	1	0
TGFB	0	0	0	0	0	0

Two genes involved in cell cycle control—TP16 and RB1—are mutually exclusive in the tissues of the central nervous system (LR 0.2, $P \leq 1.21 \times 10^{-2}$) and urinary tract (LR 0.25, $P \leq 1.19 \times 10^{-2}$). TP16 and RB1 reside along a linear pathway of cell cycle control (*i.e.*, are epistatic). The mutation of either gene suffices to release the control of cell cycle progression and cause abnormal cell growth.

Two key genes of the Wnt pathway—APC and β-catenin—are mutually exclusive in the large intestine (LR

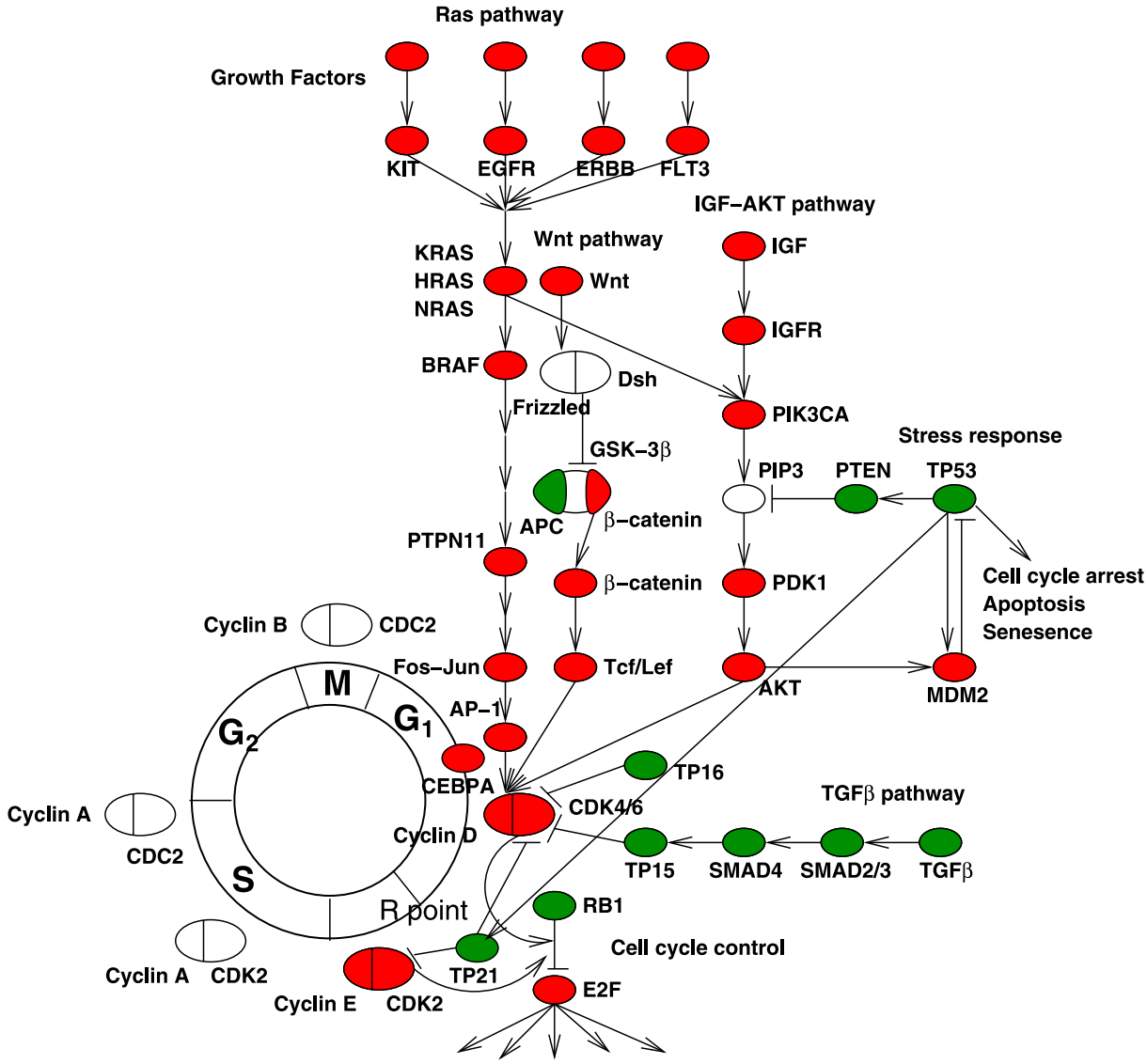

Figure 1. A simplified network of cancer pathways. Red: oncogenes. Green: tumor suppressors.

0.21, $P \leq 1.53 \times 10^{-2}$). Similar to TP16 and RB1, APC is upstream of β-catenin, regulating its degradation.

MUTUALLY EXCLUSIVE PATTERNS BETWEEN PATHWAYS AND CO-OCCURRING PATTERNS WITHIN PATHWAYS

The great majority of the combinatorial mutational patterns are consistent with the functional relations of genes as we presently understand them: Mutations of genes in different pathways co-occur, and mutations of genes in the same pathways are mutually exclusive. However, several observations violate this simple hypothesis (the diagonal entries in the top section of Table 6 and the off-diagonal entries in the bottom section of Table 6). A remarkable example is the interaction between Ras and Wnt pathways in gastrointestinal tissues. In large intestines, several pairs of genes in the two pathways comutate (APC and KRAS, β-catenin and KRAS, and APC and BRAF). For example, in 389 samples, 121 and 118 carry APC and KRAS mutations, respectively, and 45 carry both mutations (LR 1.23; $P \leq 3.47 \times 10^{-7}$). By contrast, in pancreatic mutations, the genes on these two pathways are mutually exclusive (β-catenin and KRAS, APC and KRAS). In 78 pancreatic cancers, 23 and 33 samples carry β-catenin and KRAS mutations, and only 2 carry double mutations (LR 0.21; $P \leq 6.18 \times 10^{-5}$). The distinct mutational patterns of the same set of genes in different tissues suggest the functional interactions between pathways are tissue dependent. In the large intestine, both Ras and Wnt pathways appear to be active, and mutations of genes on both pathways are frequent. In the pancreas, however, each pathway seems to be active only in specific subtypes of tissues or tumors. For instance, APC mutations are observed in the familial adenomatous polypopsis (FAP) that can give rise to pancreatic carcinomas (43), and β-catenin mutations occur in solid pseudopapillary tumors of the pancreas (44). By con-

trast, KRAS mutations are rarely observed in these tissues but are frequently detected in ductal carcinoma of the pancreas (45), where APC and β-catenin are not frequently mutated.

Several gene pairs in the Ras and IGF-AKT pathways demonstrate co-occurring mutational patterns: KRAS and PTEN (endometrium, likelihood 1.43, $P \leq 9.97 \times 10^{-3}$), BRAF and PTEN (skin; LR 1.81; $P \leq 3.56 \times 10^{-3}$), and NRAS and PTEN (ALL; LR 1.94; $P \leq 3.37 \times 10^{-2}$). However, NRAS and PTEN mutations are mutually exclusive in skin (LR 0.27; $P \leq 6.44 \times 10^{-3}$). It is argued that RAS and PTEN are functionally equivalent since both regulate the PI3-kinase activity (34). This coupling, however, cannot explain the comutations of RAS and PTEN in ALL and endometrium. The difference may be attributed to the tissue-specific different roles of RAS genes in a signal transduction pathway.

Most gene pairs in the Ras pathway have mutually exclusive mutational patterns. Yet several pairs of RAS genes demonstrate co-occurring patterns: HRAS and KRAS in soft tissue sarcomas (LR 2.49; $P \leq 1.13 \times 10^{-2}$), KRAS and NRAS in AML (LR 0.63; $P \leq 4.3 \times 10^{-3}$), and KRAS and NRAS in skin (LR 0.66; $P \leq 3.74 \times 10^{-2}$). The coexisting mutations, identical to other gain-of-function mutations of RAS genes, are observed at positions 12, 13, and 61. These anomalies may be explained by several possible causes. Even though different RAS proteins have very similar structures and functions, they may act in different pathways, exhibit dosage-specific effects, or possess subtle genetic interactions between wild-type and mutated RAS (46, 47). Alternatively, they may be artifacts from the tumor tissue mixing of single RAS mutants in the experiments.

COMBINATORIAL MUTATIONAL PATTERNS ARE HETEROGENEOUS IN DIFFERENT TISSUES

Each type of cancer possesses a specific set of combinatorial mutational patterns. We report the combinatorial patterns of each tissue in Table 5 and summarize them as follows.

The most prominent combinatorial patterns in ALL are co-occurrences among the genes in cell cycle control (TP16), stress response (TP53), IGF-AKT (PTEN), and Ras (NRAS) pathways. By contrast, the prominent combinatorial patterns in AML often contain genes involved in hematopoiesis, such as the pairs of FLT3 and CEBPA (mutually exclusive), FLT3 and NPM1 (co-occurred), and FLT3 and KIT (mutually exclusive).

TP53 and cell cycle control genes form co-occurring patterns in three types of lung cancer: adenocarcinoma, small cell carcinoma, and squamous cell carcinoma. In adenocarcinoma and squamous cell carcinoma, TP53 comutates with TP16, whereas in small cell carcinoma it comutates with RB1. Mutually exclusive patterns within the Ras pathway and co-occurring patterns among Ras, TP53, or TP16 are observed in adenocarcinoma but not in small cell carcinoma. By contrast, PTEN mutations co-occur with mutations of TP53 and RB1 in small cell carcinoma. LKB1, a serine/threonine kinase in the mTOR pathway (31), is mutually exclusive with TP53 in adenocarcinoma. This is because both LKB1 and TP53 regulate the same protein kinase (AMP-kinase) downstream of each of these tumor suppressors (31, 41).

Combinatorial patterns of TP53, TP16, and genes in Ras and Wnt pathways appear in gastrointestinal tissues such as pancreas, large intestine, and biliary tract. Most interpathway combinations are co-occurring patterns, except that the Ras-Wnt interactions in pancreas are mutually exclusive. Only one significant combinatorial pattern involved in TP16 (TP16 and KRAS are comutated) appears in the large intestine since TP16 is infrequently mutated in the large intestine (8 in 273 samples). This finding is consistent with the previous observation that TP16 is silenced by hypermethylation in colon cancer (48). Furthermore, SMAD4 (TGF-β pathway) comutates with TP16 and KRAS in pancreas and with APC and TP53 in large intestine.

In skin, intrapathway combinations of the Ras pathway genes and interpathway combinations occur with TP16, TP53, and PTEN. Most Ras pathway genes are mutually exclusive except for KRAS and NRAS mutations, which can co-occur in soft tissue tumors (sarcoma). In addition, NRAS and PTEN mutations demonstrate a mutually exclusive pattern.

In the central nervous system, both TP16 and RB1 mutations occur, but they are mutually exclusive in other tissue types. They can form co-occurring patterns with TP53 and PTEN.

THE ORDER OF MUTATIONAL EVENTS CAN BE INFERRED FROM THEIR SUBSET RELATIONSHIPS

The sequential order of gene mutations in cancer provides important information about the progression of cancer, its prevention, and treatment. A unique "path" to cancer does not exist, as many alternative sequences of mutations have been identified (*e.g.*, ref. 49). However, the order of mutations of certain genes may not be random, since the fitness of the cancer cell population may be path dependent. Following the clonal expansion model, an early mutation should be more prevalent in the clonal population than a late mutation. Therefore, if the mutation of gene A precedes the mutation of gene B, then the samples carrying B mutations should be substantially subsumed to the samples carrying A mutations. For each pair of genes, we quantify the subsumed relationships using the product of P(A mutated, B mutated) and P(B not mutated, A not mutated) and calculate the P value according to random permutations (see the Supplemental Material). **Table 7** shows 7 gene pairs that yield significant subsumed relationships ($P < 0.08$). In skin, PTEN mutations (10 samples) are completely sub-

TABLE 7. *Gene pairs yielding significant subsumed relations*

Gene				Gene mutated				
1	2	Tissue	Cosequenced	1	2	Both	Ratio	*P* value
KRAS	TP16	Pancreas	157	103	72	55	0.7639	6.82E-003
BRAF	PTEN	Skin	87	48	10	10	1	1.63E-003
TP16	PTEN	Acute lymphoblastic leukemia	31	17	8	7	0.875	3.76E-002
TP53	PTEN	Acute lymphoblastic leukemia	31	14	8	7	0.875	7.78E-003
TP53	RB1	Central nervous system	59	32	10	9	0.9	1.31E-002
APC	SMAD4	Large intestine	40	27	6	6	1	7.71E-002
KRAS	PIK3CA	Large intestine	70	35	14	10	0.7143	6.67E-002

sumed to BRAF mutations (48 samples). It is thus likely that the BRAF mutation precedes the PTEN mutation. In pancreas, TP16 mutations (72 samples) are substantially contained in KRAS mutations (103 KRAS mutation samples, 55 double mutation samples). In ALL, 7 of the 8 PTEN mutation samples are contained in TP16 (17 samples) and TP53 (14 samples) mutations, suggesting that TP16 and TP53 mutations precede PTEN mutations. In some tissues (liver, fat, muscle), TP53 can increase the level of PTEN (41). As such, TP53 inactivation would not have led to selection against PTEN. However, such selection observed in this dataset occurs in different tissue types. In the large intestine, all SMAD4 mutations (6 samples) are contained in APC mutations (27 samples), and PIK3CA mutations (14 samples) are largely contained in KRAS mutations (35 samples, 10 double mutation samples). Both results are consistent with the previous study in colon cancer (16), where APC and KRAS mutations occur at early stages of cancer progression. Finally, in the central nervous system, RB1 mutations (10 samples) are largely contained in TP53 mutations (32 samples, 9 samples of double mutations).

OBSERVED MUTATIONAL PATTERNS SUGGEST NEW COSEQUENCING TARGETS

Most previous studies are biased toward a small number of known oncogenes and tumor suppressors. To better understand the mutational landscape and combinatorial pathway/gene interactions in cancer, more genes have to be probed together. Large-scale sequencing of cancer tissues or cell lines has already been undertaken or achieved (*e.g.*, the Cancer Genome Atlas, ref. 2; the Cancer Genome Project, ref. 1). To guide the analysis of those efforts, it is possible to identify and prioritize the cosequencing targets based on the information extracted from this dataset. This "experimental design" approach can save resources and has been shown to be useful in systems biology, *e.g.* (50, 51).

We adopt two criteria to select cosequencing targets. The first criterion is to identify the gene pairs that could possibly reveal significant co-occurring or mutually exclusive patterns that are rarely sequenced together. For each tissue, we identified the gene pairs where each gene is mutated in more than 100, or 25%, of the probed samples, and both genes are cosequenced in less than 50 samples. **Table 8** lists the candidates identified by this criterion. Many gene pairs are frequently mutated but never cosequenced in any sample. For instance, in thyroid, BRAF is mutated in 35% of samples (1752 of 4936), and RET—a receptor tyrosine kinase (52)—is mutated in 36% of samples (205 of 568). Yet BRAF and RET have not been cosequenced in any sample. Other examples include TP16-PTCH (a receptor in the hedgehog signaling pathway; ref. 53) in skin, FLT3-ABL1 (a tyrosine kinase) in AML, PDGFRA-SMARCB1 (a chromatin structure regulator; ref. 54) in gastrointestinal soft-tissue sarcomas, and many more.

An alternative criterion for selecting target genes for cosequencing derives from the assumption that mutually exclusive relations are transitive. If A and B are in the same pathway, and B and C are in the same pathway, then A and C should be in the same pathway. Following the pathway hypothesis, the mutually exclusive relations of A-B and B-C are passed to A-C as well. Therefore, we can validate the transitivity of mutual exclusions by selecting the triplets A-B-C where two pairwise patterns (say A-B and B-C) are both mutually exclusive and the third pair (A-C) is cosequenced in less than 10 samples.

The only triplet passing this criterion is CEBPA-FLT3-KIT in AML. Both CEBPA-FLT3 and FLT3-KIT pairs are mutually exclusive in AML (CEBPA-FLT3: LR 0.36, $P<0.015$; FLT3-KIT: LR 0, $P<0.042$). Yet CEBPA and KIT have not been cosequenced in any AML sample. We predict they are mutually exclusive. In fact, all three of the genes are in the Ras pathway.

CONCLUSIONS AND FUTURE DIRECTIONS

The National Cancer Institute has decided to fund a pilot project designed to determine the complete nucleotide sequences in the genomes of hundreds of tumors of the same cell or tissue types so as to explore the combination of mutations that may cause these tumors. This project will undoubtedly find mutational combinations in known oncogenes, tumor-suppressor genes, and genes involved in DNA repair processes, all of which were known by previous experimentation to be causal in the formation of tumors. This project will

TABLE 8. *Candidate gene pairs for cosequencing*

	Gene		Gene probed		Gene mutated			Gene mutated		
Tissue	1	2	1	2	1	2	Cosequenced	1	2	Comutated
Thyroid	BRAF	RET	4,936	568	1752	205	0	0	0	0
Thyroid	NRAS	RET	2,628	568	174	205	0	0	0	0
Skin	BRAF	PTCH	3,603	438	1473	122	38	0	26	0
Skin	TP16	PTCH	1,191	438	272	122	0	0	0	0
Skin	NRAS	PTCH	3,144	438	606	122	38	0	26	0
Acute lymphoblastic leukemia	TP16	NOTCH1	1,123	281	377	106	0	0	0	0
Acute lymphoblastic leukemia	TP16	NRAS	1,123	1,454	377	138	31	17	8	5
Acute lymphoblastic leukemia	NOTCH1	NRAS	281	1,454	106	138	0	0	0	0
Acute myeloid leukemia	ABL1	CEBPA	558	1,989	148	199	0	0	0	0
Acute myeloid leukemia	ABL1	FLT3	558	13,436	148	3330	0	0	0	0
Acute myeloid leukemia	ABL1	GATA1	558	137	148	61	0	0	0	0
Acute myeloid leukemia	ABL1	NPM1	558	3,696	148	1169	0	0	0	0
Acute myeloid leukemia	ABL1	NRAS	558	2,665	148	336	0	0	0	0
Acute myeloid leukemia	CEBPA	NPM1	1,989	3,696	199	1169	0	0	0	0
Acute myeloid leukemia	CEBPA	NRAS	1,989	2,665	199	336	7	6	1	1
Acute myeloid leukemia	GATA1	NPM1	137	3,696	61	1169	0	0	0	0
Acute myeloid leukemia	NPM1	NRAS	3,696	2,665	1169	336	0	0	0	0
Hematopoietic/lymphoid others	JAK2	KIT	6,964	726	3743	302	0	0	0	0
Hematopoietic/lymphoid others	JAK2	NRAS	6,964	1,677	3743	185	0	0	0	0
Hematopoietic/lymphoid others	KIT	NRAS	726	1,677	302	185	0	0	0	0
Soft tissue	KIT	SMARCB1	2,736	248	975	107	0	0	0	0
Soft tissue	PDGFRA	SMARCB1	1,441	248	281	107	0	0	0	0
Breast	CDH1	PIK3CA	541	981	110	256	45	2	11	1
Stomach	B-CATENIN	KRAS	1,187	2,277	136	144	44	2	11	1
Large intestine	APC	MSH6	3,596	158	1226	48	1	1	0	0
Large intestine	B-CATENIN	PIK3CA	3,162	582	178	128	40	3	12	2
Large intestine	KRAS	MSH6	13,571	158	4207	48	4	1	0	0
Urinary tract	TP16	FGFR3	1,371	911	226	454	2	0	1	0
Urinary tract	TP16	HRAS	1,371	1,071	226	125	43	13	3	1
Urinary tract	FGFR3	HRAS	911	1,071	454	125	2	1	0	0

also undoubtedly find a large number of genetic alterations in genes and DNA sequences of no known function, some of which will be polymorphisms and some of which resulted from various genetic processes during the development of the tumor. With about 10 million known polymorphisms, the only way to identify them in this set of tumors is to also obtain the total genome sequences from normal tissue of each individual under study, a costly addition to the project. The only way to determine whether any of the changes in the genome DNA sequences of no known function truly contributes to the formation or propagation of the cancer is to test these DNA sequences or genes in model systems *in vitro* or *in vivo*. This, too, will be a costly path and may be impossible if thousands of associated genome alterations are common in a tumor and need to be tested for a growth advantage or an antiapoptotic advantage. Indeed if mutation is a random process and 5 to 6 genes in a genome are required to mutate so as to give rise to the tumor, then the tumor will surely contain tens of thousands of changes; some of which will be reduced in their number by selection, but others will be carried along by genetic drift. Therefore, only repeatedly observing a genetic alteration in independent tumors will reduce this number of events to those worth testing further, and because the number of combinations of mutations that give rise to a tumor remains unclear, the optimal number of tumors that will be needed for this complete sequencing project remains unclear. To determine if certain combinations of mutations often occur together or other combinations of mutations in genes never occur together, methods will be required to calculate the joint frequency of independent mutations expected when two genes are mutated at a certain frequency in tumors but do not cooperate or interact to form the tumor (two independent events) or both mutations work together to produce a tumor and are, therefore, found more commonly than at a random expectation. This article provides such an analysis based on existing data sets. This article also undertakes the type of analysis that is required prior to determining that we need to obtain the complete sequences of cancer genomes because we are lacking critical data in our understanding of the origins and evolution of tumors. To date, almost all of our sequencing of DNA from tumors has been accomplished by first identifying the gene or genes to be sequenced in a tumor or cell line, and this has obviously introduced an ascertainment bias in all the results presented here. This is one of the advantages to com-

plete genome sequencing, the elimination of this bias. In addition, we would like to test the hypothesis that a particular combination of gene mutations in a tumor that is found repeatedly will inform us of the properties of the tumor, the responses to therapy, the prognostic outlook, and the altered gene that should be targeted for therapy. This type of association of specific mutational patterns of oncogenes or tumor-suppressor genes with the patterns of transcription, properties of a tumor, and outcomes has been tested many times with selected tumors and genes, but this has never been done at the whole-genome level. However, some variables that clearly contribute to the properties of cancers may not be uncovered by whole-genome sequencing. These include sexual dimorphism and the expression patterns of receptors that respond to ligands that modulate transcription (ER+/− breast cancers that have very different ages of onsets, diagnostic criteria, treatments, and outcomes) and the contribution of polymorphisms that can act as modifiers of oncogenes or tumor-suppressor genes (altering the age of onset, frequency of cancers, reliance on hormones, *etc.*). It is unlikely that a genome-wide sequence will unveil a genetic modifier without further studies. It will take a thoughtful approach for this pilot project of sequencing the entire cancer genome to provide truly useful information. So what have we learned so far?

1) The cell or tissue type that will develop into a tumor determines which oncogene and tumor-suppressor gene combinations of mutations will be selected in that tumor. There is a strong tissue specificity to the pattern of mutations selected in a tumor, and this has little to do with selective expression of those genes in that tissue type. This result suggests that signal transduction pathways often act in a tissue-specific fashion, playing different roles in different tissues.

2) Certain combinations of gene mutations repeatedly occur in tissue-specific cancers but are rarely found in 100% of those cancers, suggesting other pathways to produce the same cancer types. An exception to this are some of the leukemias and lymphomas, which may have up to 100% of the same mutational patterns and therefore appear to be more simple. When combinations of mutations in three genes do occur in a tumor, they most commonly reside in genes that are located in three separate signal transduction pathways involved in the cell cycle, stress responses, and cell growth and division.

3) Some combinations of gene mutations are rarely or never found in the same tumor, even though both mutations occur at a high frequency in that tumor type. These combinations are commonly in the same signal transduction pathway, such as APC and β-catenin mutations. These types of observations confirm our present knowledge of signal transduction pathways by fulfilling the expectations for mutually exclusive mutations.

4) Combinations of mutations in genes appear to occur together in one tissue type that do not occur together in another tissue type. For example, a common co-occurrence of either KRAS and β-catenin or APC appears in tumors from the large intestine, while tumors from the pancreas have a high occurrence of KRAS and β-catenin mutations (22/78 and 33/78, respectively), but these two mutations almost never occur in the same tumor (2/78 tumors). These data demonstrate that similar signal transduction pathways act differently in different cell or tissue types.

5) Clear examples can be found of either one gene (APC) or another gene (β-catenin) in the same pathway being mutated but at different frequencies in a tumor type. Whether this finding is do to the cross-sectional size of a gene, the nature of the mutation required to inactivate or activate the gene product, or some other variable is not clear.

6) The three RAS family genes (H, K, and N) can be mutated in pairs in some tumors and exhibit mutations that are mutually exclusive in other tumors. At times, mutating two or more of these three genes in the same tumor confers a selective phenotypic advantage on the tumor. In some tissue types these genes can partially substitute for each other, whereas in other cell types they cannot substitute.

7) Signal transduction and stress response pathways are often connected, yet the co-occurring mutations between these pathways still exist. PIK3CA is regulated by RAS genes, but the comutations of KRAS and PIK3CA are observed in the large intestine, and the comutations of PTEN and RAS genes are found in skin, ALL, and endometrium. PTEN is regulated by TP53 in some tissue types (41), but their comutations are found in six tissue types. This finding suggests that the connectivity alone does not suffice to determine the mutational patterns. Tissue-specific gene expressions and effects on other genes may also play roles in selection.

8) The patterns of mutations in cell lines when compared to tumors have some clear differences (increased p53 and ARF mutations, *etc.*), but many similarities between the cells growing in these very different environments remain.

9) It has been possible to examine the frequencies of different mutations in a tumor group and the preferred coexistence of two mutations in that group of tumors to discern an order of the mutations that were selected during the development of the tumor. The detected ordering of these random mutations does imply that this process indeed plays a role in tumor formation. Some types of mutations may be "gatekeepers," or the first mutation could raise the mutation rate for subsequent mutations. If this concept is correct and common, we will need to develop models to test these ideas.

The results of this analysis point out both the limitations and the advantages in carrying out whole-genome sequencing of cancer genomes *vs.* selected genome sequencing. Clearly we have learned some things from our progress to date. It will be interesting to see what the NCI pilot project adds to this information. FJ

REFERENCES

1. The Cancer Genome Project. Wellcome Trust Sanger Institute. http://www.sanger.ac.uk/genetics/CGP/

2. The Cancer Genome Atlas. National Cancer Institute. http://cancergenome.nih.gov/
3. Forbes, S., Clements, J., Dawson, E., Bamford, S., Webb, T., Dogan, A., Flanagan, A., Teague, J., Wooster, R., Futreal, P. A., and Stratton, M. R. (2006) COSMIC 2005. *Br. J. Cancer* **94,** 318–322
4. Cancer Cell Line Project. Wellcome Trust Sanger Institute. http://www.sanger.ac.uk/genetics/CGP/CellLines/
5. Lyman S. D, James L, Vanden Bos T, de Vries P, Brasei K, Gliniak B, Hollingsworth L. T, Picha K. S, McKenna H. J, Splett R. R., Fletcher, F. A., Maraskovsky, E., Farrah, T., Foxworthe, D., Williams, D. E., and Beckmann, M. P. (1993) Molecular cloning of a ligand for the flt3/flk-2 tyrosine kinase receptor: a proliferative factor for primitive hematopoietic cells. *Cell* **75**(6), 1157–1167
6. Kiyoi, H., Naoe, T. (2006) FLT3 mutations in acute myeloid leukemia. *Methods Mol. Med.* **125,** 189–197
7. Matthews, W., Jordan, C. T., Wiegand, G. W., Pardoll, D., and Lemischka, I. R. (1991) A receptor tyrosine kinase specific to hematopoietic stem and progenitor cell-enriched populations. *Cell* **65,** 1143–1152
8. Bruening, W., Bardeesy, N., Silverman, B. L., Cohn, R. A., Machin, G. A., Aronson, A. J., Housman, D., and Pelletier, J. (1992) Germline intronic and exonic mutations in the Wilms' tumour gene (WT1) affecting urogenital development. *Nat. Genet.* **1,** 144–148
9. Hynes, N. E., and Lane, H. A. (2005) ERBB receptors and cancer: the complexity of targeted inhibitors. *Nat. Rev. Cancer* **5,** 341–354
10. Hirsch F. R, Varella-Garcia M, Bunn P. A, Jr., Di Maria, M. V., Veve, R., Bremnes, R. M., Baron, A. E., Zeng, C., and Franklin, W. A. (2003). Epidermal growth factor receptor in non-small-cell lung carcinomas: correlation between gene copy number and protein expression and impact on prognosis. *J. Clin. Oncol.* **21**, 3798–3807
11. Sonobe, M., Manabe, T., Wada, H., and Tanaka, F. (2005) Mutations in the epidermal growth factor receptor gene are linked to smoking-independent, lung adenocarcinoma. *Br. J. Cancer* **93,** 355–363
12. Nakao, M., Yokota, S., Iwai, T., Kaneko, H., Horlike, S., Kashima, K., Sonoda, Y., Fujimoto, T., and Misawa, S. (1996) Internal tandem duplication of the flt3 gene found in acute myeloid leukemia. *Leukemia* **10,** 1911–1918
13. Ciesielski, M. J., and Fenstermaker, R. A. (2000) Oncogeic epidermal growth factor receptor mutants with tandem duplication: gene structure and effects on receptor function. *Oncogene* **19,** 810–820
14. Ji, H., Zhao, X., Yuza, Y., Shimamura, T., Li, D., Protopopov, A., Jung, B. L., McNamara, K., Xia, H., Glatt, K. A., Thomas, R. K., Sasaki, H., Horner, J. W., Eck, M., Mitchell, A., Sun, Y., Al-Hashem, R., Bronson, R. T., Rabindran, S. K., Discafani, C. M., Maher, E., Shapiro, G. I., Meyerson, M., and Wong, K. K. (2006) Epidermal growth factor receptor variant III mutations in lung tumorigenesis and sensitivity to tyrosine kinase inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 7817–7822
15. Rubinfeld, B., Robbins, P., El-Gamil, M., Albert, I., Porfiri, E., and Polakis, P. (1997) Stabilization of beta-catenin by genetic defects in melanoma cell lines. *Science* **275,** 1790–1792
16. Vogelstein, B., Fearon, E. R., Hamilton, S. R., Kern, S. E., Preisinger, A. C., Leppert, M., Nakamura, Y., White, R., Smits, A. M., and Bos, J. L. (1988) Genetic alterations during colorectal-tumor development. *N. Engl. J. Med.* **319,** 525–532
17. Cahill, D. P., Kinzler, K. W., Vogelstein, B., and Lengauer, C. (1999) Genetic instability and darwinian selection in tumours. *Trends Cell Biol.* **9,** M57–60
18. Serrano, M., Hannon, G. J., and Beach, D. (1993) A new regulatory motif in cell-cycle control causing specific inhibition of cyclin D/CDK4. *Nature* **366,** 704–707
19. Lukas, J., Parry, D., Aagaard, L., Mann, D. J., Bartkova, J., Strauss, M., Peters, G., and Bartek, J. (1995) Retinoblastoma-protein-dependent cell-cycle inhibition by the tumour suppressor p16. *Nature* **375,** 503–506
20. Blais, A., and Dynlacht, B. D. (2004) Hitting their targets: an emerging picture of E2F and cell cycle control. *Curr. Opin. Genet. Dev.* **14,** 527–532
21. Rubin, B. P., Heinrich, M. C., and Corless, C. L. (2007) Gastrointestinal stromal tumour. *Lancet* **369,** 1731–1741
22. McCubrey, J. A., May, W. S., Duronio, V., and Mufson, A. (2000) Serine/threonine phosphorylation in cytokine signal transduction. *Leukemia* **14,** 9–21
23. Kahn, S., Yamamoto, F., Almoguera, C., Winter, E., Forrester, K., Jordano, J., and Perucho, M. (1987) The c-K-ras gene and human cancer (review). *Anticancer Res.* **7(4A)**, 639–652
24. Barbacid, M. (1987) Ras genes. *Annu. Rev. Biochem.* **56,** 779–827
25. Hancock, J. F. (2003) Ras proteins: different signals from different locations. *Nat. Rev. Mol. Cell. Biol.* **4,** 373–385
26. Kolch, W. (2000) Meaningful relationships: the regulation of the Ras/Raf/MEK/ERK pathway by protein interactions. *Biochem. J.* **351,** 289–305
27. Peyssonnaux, C., and Eychène, A. (2001) The Raf/MEK/ERK pathway: new concepts of activation. *Biol. Cell* **93,** 53–62
28. Chan, R. J., and Feng, G. S. (2007) PTPN11 is the first identified proto-oncogene that encodes a tyrosine phosphatase. *Blood* **109,** 862–867
29. Wang, H., Iakova, P., Wilde, M., Welm, A., Goode, T., Roesler, W. J., and Timchenko, N. A. (2001) CEBPalpha arrests cell proliferation through direct inhibition of Cdk2 and Cdk4. *Mol. Cell* **8,** 817–828
30. Lustig, B., and Behrens, J. (2003) The Wnt signaling pathway and its role in tumor development. *J. Cancer Res. Clin. Oncol.* **129,** 199–221
31. Levine, A. J., Feng, Z., Mak, T. W., You, H., and Jin, S. (2006) Coordination and communication between the p53 and IGF-1-AKT-TOR signal transduction pathways. *Genes Dev.* **20,** 267–275
32. Maehama, T., and Dixon, J. E. (1998) The tumor suppressor PTEN/MMAC1 dephosphorylates the lipid second messenger phosphatidylinositol 3,4,5-trisphosphate. *J. Biol. Chem.* **273,** 13375–13378
33. Rodriguez-Viciana, P., Warne, P. H., Khwaja, A., Marte, B. M., Pappin, D., Das, P., Waterfield, M. D., Ridley, A., and Downward, J. (1997) Role of phosphoinositide 3-OH kinase in cell transformation and control of the actin cytoskeleton by Ras. *Cell* **89,** 457–467
34. Tsao, H., Zhang, X., Fowlkes, K., and Haluska, F. G. (2000) Relative reciprocity of NRAS and PTEN/MMAC1 alterations in cutaneous melanoma cell lines. *Cancer Res.* **60,** 1800–1804
35. Nakao, A., Imamura, T., Souchelnytskyi, S., Kawabata, M., Ishisaki, A., Oeda, E., Tamaki, K., Hanai, J., Heldin, C. H., Miyazono, K., and ten Dijke, P. (1997) TGF-beta receptor-mediated signalling through Smad2, Smad3 and Smad4. *EMBO J.* **16,** 5353–5362
36. Liu, F., Pouponnot, C., and Massague, J. (1997) Dual role of the Smad4/DPC4 tumor suppressor in TGFbeta-inducible transcriptional complexes. *Genes Dev.* **11,** 3157–3167
37. Peng, B., Fleming, J. B., Breslin, T., Grau, A. M., Fojioka, S., Abbruzzese, J. L., Evans, D. B., Ayers, D., Wathen, K., Wu, T., Robertson, K. D., and Chiao, P. J. (2002) Suppression of tumorigenesis and induction of p15(ink4b) by Smad4/DPC4 in human pancreatic cancer cells. *Clin. Cancer Res.* **8,** 3628–3638
38. Levine, A. J. (1997) p53, the cellular gatekeeper for growth and division. *Cell* **88,** 323–331
39. Vogelstein, B., Lane, D., and Levine, A. J. (2000) Surfing the p53 network. *Nature* **408,** 307–310
40. El-Deiry, W. S., Harper, J. W., O'Connor, P. M., Velculescu, V. E., Canman, C. E., Jackman, J., Pietenpol, J. A., Burrell, M., Hill, D. E., Wang, Y., Wiman, K. G., Mercer, W. E., Kastan, M. B., Kohn, K. W., Elledge, S. J., Kinzler, K. W., and Vogelstein, B. (1994) WAF1/CIP1 is induced in p53-mediated G1 arrest and apoptosis. *Cancer Res.* **54(5)**, 1169–1174
41. Feng, Z., Hu, W., de Stanchina, E., Teresky, A. K., Jin, S., Lowe, S., Levine, A. J. (2007) The regulation of AMPK beta1, TSC2, and PTEN expression by p53: stress, cell and tissue specificity, and the role of these gene products in modulating the IGF-1-AKT-mTOR pathways. *Cancer Res.* **67,** 3043–3053
42. Christiansen, D. H., Andersen, M. K., Desta, F., Pedersen-Bjergaard, J. (2005) Mutations of genes in the receptor tyrosine kinase (RTK)/RAS-BRAF signal transduction pathway in therapy-related myelodysplasia and acute myeloid leukemia. *Leukemia* **19,** 2232–2240
43. Toyooka, M., Konishi, M., Kikuchi-Yanoshita, R., Iwama, T., and Miyaki, M. (1995) Somatic mutations of the adenomatous polyposis coli gene in gastroduodenal tumors from patients with familial adenomatous polyposis. *Cancer Res.* **55,** 3165–3170

44. Abraham, S. C., Klimstra, D. S., Wilentz, R. E., Yeo, C. J., Conlon, K., Brennan, M., Cameron, J. L., Wu, T. T., and Hruban, R. H. (2002) Solid-pseudopapillary tumors of the pancreas are genetically distinct from pancreatic ductal adenocarcinomas and almost always harbor beta-catenin mutations. *Am. J. Pathol.* **160,** 1361–1369
45. Ku, J. L., Yoon, K. A., Kim, W. H., Jang, Y., Suh, K. S., Kim, S. W., Park, Y. H., and Park, J. G. (2002) Establishment and characterization of four human pancreatic carcinoma cell lines. Genetic alterations in the TGFBR2 gene but not in the MADH4 gene. *Cell Tissue Res.* **308,** 205–214
46. To, M. D, Perez-Losada, J., Mao, J. H., Hsu, J., Jacks, T., and Balmain, A. (2006) A functional switch from lung cancer resistance to susceptibility at the *Pas1* locus *Kras2*^{LA2} mice. *Nat. Genet.* **38**(8), 926–930
47. Diaz, R., Lue, J., Mathews, J., Yoon, A., Ahn, D., Garcia-Espana, A., Leonardi, P., Vargas, M. P., and Pellicer, A. (2005) Inhibition of Ras oncogenic activity by Ras protooncogenes. *Int. J. Cancer* **113,** 241–248
48. Herman, J. G., Merlo, A., Mao, L., Lapidus, R. G., Issa, J. P., Davidson, K. E., Sidransky, D., and Baylin, S. B. (1995) Inactivation of the CDKN2/p16/MTS1 gene is frequently associated with aberrant DNA methylation in all common human cancers. *Cancer Res.* **55,** 4525–4530
49. Barrett, M. T., Sanchez, C. A., Prevo, L. J., Wong, D. J., Galipeau, P. C., Paulson, T. G., Rabinovitch, P. S., and Reid, B. J. (1999) Evolution of neoplastic cell lineages in Barrett oesophagus. *Nat. Genetics* **22,** 106–109
50. King, R. D., Whelan, K. E., Jones, F. M., Reiser, P. G., Bryant, C. H., Muggleton, S. H., Kell, D. B., and Oliver, S. G. (2004) Functional genomic hypothesis generation and experimentation by a robot scientist. *Nature* **427,** 247–252
51. Yeang, C. H., Mak, H. C., McCuine, S., Workman, C., Jaakkola, T., and Ideker, T. (2005) Validation and refinement of gene-regulatory pathways on a network of physical interactions. *Genome Biol.* **6,** R62
52. Van der Geer, P., Hunter, T., and Lindberg, R. A. (1994) Receptor protein-tyrosine kinases and their signal transduction pathways. *Annu. Rev. Cell Biol.* **10,** 251–337
53. Hahn, H., Christiansen, J., Wicking, C., Zaphiropoulos, P. G., Chidambaram, A., Gerrard, B., Vorechovsky, I., Bale, A. E., Toftgard, R., Dean, M., and Wainwright, B. (1996) A mammalian patched homolog is expressed in target tissues of sonic hedgehog and maps to a region associated with developmental abnormalities. *J. Biol. Chem.* **271,** 12125–12128
54. Hirschhorn J. N, Brown S. A, Clark C. D, and Winston, F. (1992) Evidence that SNF2/SWI2 and SNF5 activate transcription in yeast by altering chromatin structure. *Genes Dev.* **6**(12A), 2288–2298

Received for publication March 3, 2008.
Accepted for publication March 20, 2008.