ARTICLE

Quantifying Histological Features of Cancer Biospecimens for Biobanking Quality Assurance Using Automated Morphometric Pattern Recognition Image Analysis Algorithms

Joshua D. Webster,[1] Eleanor R. Simpson,[2] Aleksandra M. Michalowski,[1] Shelley B. Hoover,[1] and R. Mark Simpson[1,]*

[1]Laboratory of Cancer Biology and Genetics and [2]Pediatric Oncology Branch, Center for Cancer Research, National Cancer Institute, Bethesda, Maryland 20892, USA

Biorepository-supported translational research depends on high-quality, well-annotated specimens. Histopathology assessment contributes insight into how representative lesions are for research objectives. Feasibility of documenting histological proportions of tumor and stroma was studied in an effort to enhance information regarding biorepository tissue heterogeneity. Using commercially available software, unique spatial-spectral algorithms were developed for applying automated pattern recognition morphometric image analysis to quantify histologic tumor and nontumor tissue areas in biospecimen tissue sections. Measurements were acquired successfully for 75/75 (100%) lymphomas, 76/77 (98.7%) osteosarcomas, and 60/70 (85.7%) melanomas. The percentage of tissue area occupied by tumor varied among patients and tumor types and was distributed around medians of 94% [interquartile range (IQR)=14%] for lymphomas, 84% for melanomas (IQR=24%), and 39% for osteosarcomas (IQR=44%). Within-patient comparisons from a subset, including multiple individual patient specimens, revealed ≤12% median coefficient of variation (CV) for lymphomas and melanomas. Phenotypic heterogeneity of osteosarcomas resulted in 33% median CV. Uniformly applied, tumor-specific pattern recognition software permits automated tissue-feature quantification. Furthermore, dispersion analyses of area measurements across collections, as well as of multiple specimens from individual patients, support using limited tissue slices to gauge features for some tumor types. Quantitative image analysis automation is anticipated to minimize variability associated with routine biorepository pathologic evaluations and enhance biomarker discovery by helping to guide the selection of study-appropriate specimens.

KEY WORDS: biological specimens bank, standards, biomarkers, isolation and purification, morphometry, quantitative pathology, translational research, methods

INTRODUCTION

Advancements in individualized patient care are being sought through the identification and development of disease-relevant biomarkers. In the case of oncology, cancer tissue resource facilities play an essential function in facilitating biomarker discovery and translational research through critical roles in the identification and validation of novel diagnostic and prognostic tests, therapeutic targets, and identification of markers, which can be used to predict a therapeutic response. Biomarker discovery requires the availability of high-quality tissue specimens for macromolecule extraction and downstream analyses. The need for well-characterized tissues from diseased patients has led to recent research into biospecimens and their collation in and distribution from biospecimen tissue resource facilities (http://cahub.cancer.gov/).[1–4] Typically, the entirety of a banked specimen is extracted to yield material for most biomarker discovery research. As tumor specimens are composed of heterogeneous populations of cells (including viable and necrotic neoplastic cells), stromal elements (including fibroblasts, vasculature, nervous supply, infiltrating leukocytes, and ECM), and adjacent normal tissues, the various proportions of the whole array of cells contribute to extracted macromolecules for molecular analyses.[2,5] Moreover, tissues must contain suitable proportions of neoplastic tissue to enable collection of disease-relevant molecules for analyses.[1,6]

Pathology assessment provides insight into the contribution of the various cellular components of banked tissue

*ADDRESS CORRESPONDENCE TO: R. Mark Simpson, Laboratory of Cancer Biology and Genetics, Center for Cancer Research, NCI, NIH, Building 37, Room 2000, Bethesda, MD 20892, USA (E-mail: ms43b@nih.gov).

There are no conflicts of interest.

specimens, such as neoplastic cells and stroma, and is essential for understanding how representative a lesion in a specimen is for intended study applications.[1,3,6] Too frequently, inattention to the histological features of research specimens contributes to misconception of molecular findings; discrepant results can occur as a result of tissue heterogeneity, confounded by the presence of normal tissue and reparative reactions.[1] Understanding specimen characteristics is critical for translating clinically relevant findings.[2] Previous morphological evaluations of biorepository specimens revealed that up to 17% of specimens may lack the intended tumor, making them unsuitable for molecular discovery.[6,7] Similarly, pathologist assessments have been used to limit molecular analyses to specimens estimated to have >65% tumor.[6] Through increased appreciation of relevant pathologic features in biospecimens, greater contextual understanding for derived macromolecules may be forthcoming, helping to augment the probability of translational success.[4] The finding that 65% of the tumor biospecimens acquired for The Cancer Genome Atlas research purposes was of inadequate quality led the NCI's The Cancer Human Biobank (caHUB) biorepository network to address the national shortage of high-quality cancer-associated blood and tissue specimens for use in translational biomedical research (http://cahub.cancer.gov/). caHUB tissue collection guidelines include criteria for histopathologic confirmation of specimens' diagnoses and institute a general minimum criterion that 70% of each banked tissue specimen should contain viable tumor, and the remaining 30% of the tissue can be composed of other cell types.

Precisely how the histologic composition of biorepository specimens should be determined is at the discretion of the biorepository or contributor. Histologic assessments for composition of biospecimens are generally based on pathologists' visual estimations from histologic tissue sections considered to be representative. Such estimates are subjective and can be associated with intra- and inter-observer variation, especially if assessments are performed by multiple pathologists at multiple institutions. Computer-assisted histological image analysis can be used to obtain estimates of the proportion of tissue features within cancer biospecimens. This has the potential to enable biobanks to provide information about the various contributing cell types likely contained in each specimen. Furthermore, automated image analysis minimizes subjectivity and intra- and inter-observer variation associated with a pathologist's visual estimates of tissue proportions.[8] Such pattern recognition analyses of digitized image data files are finding their way into automated scoring for immunohistochemistry.[9] With more adequately characterized biospecimens, investigators can choose study-suitable tissues. Therefore, in this study, we sought to create unique image analysis algorithms, using commercially available histologic pattern recognition software to quantify the proportion of tumor and nontumor tissues in biorepository tissue specimens. The objective of this study was to determine the feasibility for documenting histological proportions of tumor and nontumor tissues using these algorithms to annotate divergent specimen characteristics for biobank investigators as a means to minimize the subjectivity and intra- and inter-observer variability associated with visual estimations performed over time.

MATERIALS AND METHODS

Standardized, automated methods suitable for augmenting a biobank pathology quality-assurance program were sought by using computer-assisted genetic image evaluation pattern recognition software analyses (Genie, Aperio Technologies, Vista, CA, USA). Genie, which was developed originally for satellite photographic image analysis by Los Alamos National Laboratory, U.S. Department of Energy, uses an iterative genetic process to identify unique spatial-spectral features of user-defined tissue-class examples that can discriminate tissue features within histologic sections. In this study, Genie algorithms were developed from and applied to images obtained from routinely processed tissue sections of spontaneous neoplasms, collected prior to the onset of therapy, from dogs admitted to academic veterinary teaching hospitals throughout the United States as part of the Pfizer Canine Comparative Oncology & Genomics Consortium (CCOGC; Rockville, MD, USA) biospecimen repository. Protocols for tissue collection, processing, and preservation are available at the biorepository website (http://www.ccogc.net/collection.html). Analogous to caHUB, the CCOGC biorepository includes frozen and fixed tumor specimens, blood, serum, urine, and genomic DNA and RNA samples, as well as cancer-adjacent normal tissue specimens from canine patients with spontaneous neoplastic diseases. These are used in informative translational and comparative cancer biology research, which has been reviewed recently.[10]

Routinely processed tissue sections from formalin-fixed, paraffin-embedded tissues were stained with H&E (Histoserv, Germantown, MD, USA). Glass slides were optically scanned using a ScanScope XT digital slide scanner (Aperio Technologies) to create whole-slide digital planar image files at 0.5 μm/pixel resolution. The original diagnosis provided to the repository was corroborated by microscopic histological assessment. If the diagnosis could not be confirmed or if tumor tissue was not identified in the provided tissue section, the specimen was excluded from further analysis. Specimens analyzed included 75 lympho-

mas from 49 patients, 77 osteosarcomas from 43 patients, and 70 melanomas from 45 patients.

Unique algorithms for each cancer type (lymphoma, osteosarcoma, and melanoma) were developed independently. Algorithms for discriminating tissue feature area measurements within study slide images were created using training slide sets, which contained representative examples of each tissue class for each cancer type (Table 1). Each algorithm was developed to segment and quantify areas of tumor, based on the unique histological features of each cancer type versus areas of non-neoplastic tissue, which included stroma, necrosis, and normal adjacent tissues. Representative areas of each tissue class to be segmented were selected to create digital montages in an approach similar to a previous study[9] and following parameters similar to those featured in caHUB collection guidelines (http://cahub.cancer.gov/). For training, the software applied an iterative genetic learning process to discriminate identified tissue classes based on unique spatial-spectral features for each tumor type[11] and reviewed montages independently as unknowns to determine sensitivities and specificities for each tissue class within the tumor training sets (Table 1). Tumor-specific morphometric algorithms were subsequently run on image data files of biobank specimens. Morphological acceptability of image area segmentation was judged by visual inspection of pseudo-color image mark-up files. Based on these examinations, relative heterogeneity observed within melanomas and osteosarcomas necessitated development of a second pattern recognition algorithm for each to accommodate several specimens with dramatic image segmentation inaccuracy. Training sets for second algorithms comprised those specimens poorly segmented by the original algorithm. Combined results obtained for the collection by running both algorithms included all except one osteosarcoma and 10 melanoma tissue sections, resulting in the inability to tabulate one osteosarcoma patient and four melanoma patients. Consequently, a total of five algorithms was used for image analysis of the three tumor types.

Frequency distributions of the tumor percentage were estimated using kernel density estimates,[12] implemented in The R Project for Statistical Computing R *stats* package (http://www.R-project.org). The normality, skewness, and kurtosis of the tumor percentage distributions were tested

TABLE 1

Tissue Feature Classes and Genie Training Results Developed for Tumor Area Quantification in Biorepository Specimens Using Tumor-Specific Algorithms

Tumor algorithm	Tissue class	Sensitivity (%)	Specificity (%)	Mean training accuracy (%)
Lymphoma	**Tumor**[a]	99.98	99.99	99.98
	Stroma (muscle, connective tissue)	99.97	100	
	Glass[b]	100	99.89	
Osteosarcoma 1	**Tumor**	99.54	97.45	99.22
	Stroma (muscle, connective tissue, bone, acellular osteoid, hemorrhage, necrosis)	98.25	99.77	
	Glass	99.86	99.61	
Osteosarcoma 2	**Tumor**	97.61	98.26	98.33
	Stroma (muscle, connective tissue, bone, acellular osteoid, hemorrhage, necrosis)	97.66	96.92	
	Glass	99.72	99.43	
Melanoma 1	**Unpigmented tumor**	99.08	98.9	99.29
	Pigmented tumor	99.7	99.05	
	Stroma (muscle, connective tissue, epithelium, hemorrhage, necrosis)	98.84	99.11	
	Glass	99.53	99.46	
Melanoma 2	**Combined pigmented and unpigmented tumor**	99.14	95.28	98.51
	Stroma (muscle, connective tissue, epithelium, hemorrhage, necrosis)	96.44	99.38	
	Glass	99.94	99.01	

[a]Boldface font represents primary tissue classes used for training and testing with various features included within classes listed in parentheses. [b]Glass represents the background of the microscope slide.

with the Shapiro-Wilk[13] (R *stats* package), D'Agostino,[14] and Anscombe-Glynn[15] (R *moments* package) tests, respectively. Box-plot statistics were defined as in Tukey[16] and generated with R *graphics* package. R version 2.10 was used for the above analyses (http://www.R-project.org).[17]

RESULTS

Unique cancer-specific histologic pattern recognition algorithms for lymphoma, osteosarcoma, and melanoma were developed. Refinements, achieved through a series of reoptimizations of user-defined examples of each tissue class identified in training sets, resulted in mean training accuracies of >98%, with >95% sensitivities and specificities for each tissue class across algorithms developed during training for all tumor types (Table 1). When applied to individual biobank specimens (Fig. 1), area and percentage estimates of tumor and stroma histology were obtained for 75/75 (100%) lymphomas, 76/77 (98.7%) osteosarcomas, and 60/70 (85.7%) melanomas (Table 2).

The percentage of total tissue specimen area occupied by tumor for each patient was used to examine the distribution of the proportion of tumor, compared with nontumor, among biobank specimens for each tumor phenotype (Fig. 2*A* and Table 2). The median tumor content in lymphoma patients reached 94% and varied the least among the three tumor phenotypes (IQR=14%). The percentages of tumor tissue areas in melanoma patients were distributed around the median of 84% with a higher spread than in the lymphoma patients (IQR=24%). The range of tumor tissue content in the osteosarcoma patients was the greatest (IQR=44%), with the median portion of tissue area occupied by tumor reaching 39%. As indicated by the frequency distributions, lymphoma and melanoma tumor area content distributions exhibited negative skewness ($P<0.05$, D'Agostino test) and highly significant departure from normally distributed data ($P<0.001$, Shapiro-Wilk test). The distribution of tumor area percentages in osteosarcoma patients deviated less significantly from normality ($P<0.05$, Shapiro-Wilk test); however, this could not be explained by deviation from symmetry ($P=0.3$, D'Agostino test) or excessive kurtosis ($P=0.06$, Anscombe-Glynn test).

Pathology-based biorepository quality-assurance programs rely on the assumption that evaluations of single tissue sections from the biobank specimen can be used as a representative surrogate measure for the distribution of tissue components present in the remaining sample, contained in the tissue biobank. Further insight into how well area measurements from single tissue slices may represent a given banked tumor was gained by comparing morphometry on cases for which institutions had submitted multiple biopsy specimens from individual patients. These within-patient comparisons were limited by the number of available cases that included two or more specimens from individual patients. Single tissue sections of multiple tumor specimens were available from 13 lymphoma, 22 osteosarcoma, and 12 melanoma patients (lesion multiplicities collected during the same patient visit). Median CV among specimens from individual patients was 2% for lymphomas, 12% for melanomas, and 33% for osteosarcomas (Fig. 2*B*).

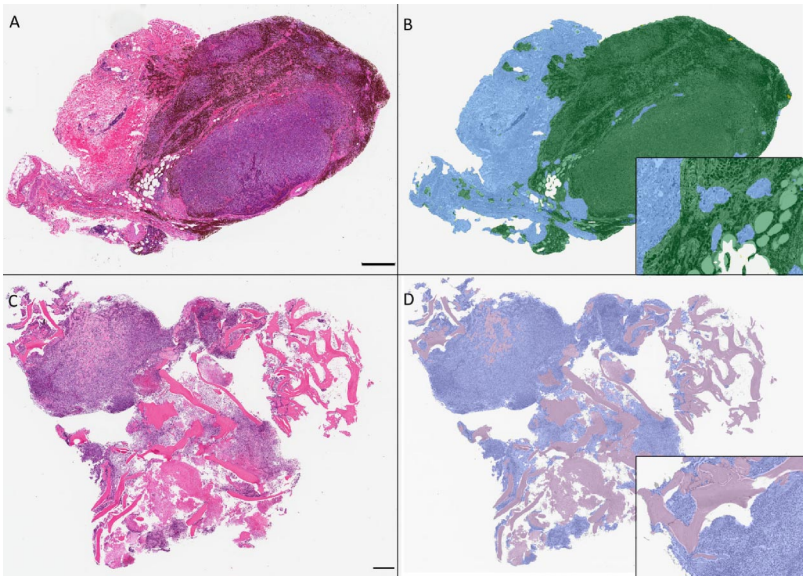

FIGURE 1

Automated pattern recognition image analysis approach for quantifying tissue features in biobank tissue specimens as part of pathology quality assurance. Unique algorithms were developed for each tumor type. (*A*) Sub-gross photomicrograph of an optically scanned melanoma tissue section image used for morphometric analysis is displayed as a representative example. H&E stain. (*B*) Tissue section shown in *A* after image analysis, displayed with pseudo-color mark-up, revealing tissue areas segmented as cancer (green, 68.6%) and stroma (blue, 31.4%). A postprocessing adjustment to the pseudo-color mark-up image was made to adjust the background to white. (*C*) Sub-gross photomicrograph of an optically scanned osteosarcoma tissue section image used for morphometric analysis is displayed as a representative example. H&E stain. (*D*) Tissue section shown in *C* after image analysis, displayed with pseudo-color mark-up, revealing tissue areas segmented as cancer (blue, 55.6%) and stroma (gray, 44.6%). Insets (*B* and *D*) demonstrate the higher magnification demonstrations of the algorithms' ability to segment tumor and stroma. Original bars = 500 μm.

TABLE 2

Complete Pathology Quality-Assurance Data Output Specifying Diagnosis and Morphometric Tissue Feature Quantification from Computer-Assisted Pattern Recognition Image Analysis of Digitally Scanned Tissue Sections from 49 Lymphoma, 43 Osteosarcoma, and 45 Melanoma Patients

Patient number[a]	Diagnosis[b]	Tissue area (mm^2)[c]	Tumor area (mm^2)	Stroma area (mm^2)	Percent tumor	Percent stroma
1	Lymphoma	40.40	39.36	1.04	97.42	2.58
2	Lymphoma	35.32	34.45	0.87	97.54	2.46
3	Lymphoma	23.31	23.22	0.09	99.63	0.37
4	Lymphoma	71.65	67.81	3.84	94.64	5.36
5	Lymphoma	37.51	36.84	0.67	98.22	1.78
6	Lymphoma	7.57	6.99	0.58	92.38	7.62
7	Lymphoma	39.23	25.48	13.75	64.95	35.05
8	Lymphoma	27.31	24.86	2.45	91.04	8.96
9A	Lymphoma	51.91	48.88	3.03	94.16	5.84
9B	Lymphoma	52.21	49.71	2.50	95.21	4.79
10	Lymphoma	18.73	14.53	4.20	77.58	22.42
11	Lymphoma	53.76	49.30	4.46	91.70	8.30
12	Lymphoma	54.72	51.46	3.26	94.05	5.95
13	Lymphoma	19.39	15.65	3.74	80.74	19.27
14	Lymphoma	30.59	30.04	0.55	98.20	1.80
15A	Lymphoma	103.35	100.57	2.78	97.31	2.69
15B	Lymphoma	94.79	93.10	1.69	98.22	1.78
16A	Lymphoma	16.41	14.97	1.44	91.23	8.77
16B	Lymphoma	22.96	21.29	1.67	92.72	7.28
16C	Lymphoma	17.45	15.35	2.10	87.99	12.01
16D	Lymphoma	28.58	24.94	3.64	87.26	12.74
17A	Lymphoma	18.65	16.83	1.82	90.25	9.75
17B	Lymphoma	15.15	14.23	0.92	93.95	6.05
18A	Lymphoma	20.52	14.45	6.07	70.44	29.56
18B	Lymphoma	7.56	6.34	1.22	83.82	16.18
18C	Lymphoma	19.76	18.30	1.46	92.60	7.40
19A	Lymphoma	18.35	18.30	0.05	99.71	0.29
19B	Lymphoma	49.74	48.98	0.76	98.47	1.53
19C	Lymphoma	38.80	38.70	0.10	99.74	0.26
20A	Lymphoma	10.94	9.83	1.11	89.83	10.17
20B	Lymphoma	25.32	24.52	0.80	96.85	3.15
20C	Lymphoma	28.86	28.83	0.03	99.88	0.12
21A	Lymphoma	18.52	13.65	4.87	73.70	26.30
21B	Lymphoma	27.57	24.42	3.15	88.58	11.42
21C	Lymphoma	23.61	18.66	4.95	79.04	20.96
22A	Lymphoma	46.28	45.75	0.53	98.86	1.14
22B	Lymphoma	14.92	14.46	0.46	96.90	3.10
22C	Lymphoma	22.78	21.78	1.00	95.62	4.38
23A	Lymphoma	12.10	12.09	0.01	99.91	0.09
23B	Lymphoma	9.70	9.68	0.02	99.81	0.19
23C	Lymphoma	8.73	8.70	0.03	99.67	0.33
24A	Lymphoma	34.35	32.64	1.71	95.02	4.98
24B	Lymphoma	70.18	66.36	3.82	94.56	5.44
24C	Lymphoma	88.01	80.54	7.47	91.52	8.48
25A	Lymphoma	27.92	27.86	0.06	99.80	0.20

TABLE 2

Continued

Patient number[a]	Diagnosis[b]	Tissue area (mm^2)[c]	Tumor area (mm^2)	Stroma area (mm^2)	Percent tumor	Percent stroma
25B	Lymphoma	19.53	19.44	0.09	99.56	0.44
25C	Lymphoma	20.12	19.52	0.60	97.03	2.97
26	Lymphoma	37.73	34.73	3.00	92.05	7.95
27	Lymphoma	44.09	41.53	2.56	94.20	5.80
28	Lymphoma	27.19	15.73	11.46	57.86	42.14
29	Lymphoma	45.80	37.36	8.44	81.57	18.43
30	Lymphoma	40.84	34.25	6.59	83.88	16.12
31	Lymphoma	12.16	11.63	0.53	95.62	4.38
32	Lymphoma	43.79	43.64	0.15	99.65	0.35
33	Lymphoma	74.61	71.48	3.13	95.80	4.20
34	Lymphoma	101.58	96.10	5.48	94.60	5.40
35	Lymphoma	98.66	95.80	2.86	97.10	2.90
36	Lymphoma	22.17	21.80	0.37	98.34	1.66
37	Lymphoma	103.28	97.08	6.20	93.99	6.01
38	Lymphoma	56.91	43.49	13.42	76.42	23.58
39	Lymphoma	70.62	65.18	5.44	92.29	7.71
40	Lymphoma	69.84	58.22	11.62	83.36	16.64
41	Lymphoma	18.94	14.34	4.60	75.70	24.30
42	Lymphoma	119.06	107.68	11.38	90.44	9.56
43	Lymphoma	74.40	73.63	0.77	98.96	1.04
44	Lymphoma	52.31	49.95	2.36	95.48	4.52
45A	Lymphoma	149.28	144.00	5.28	96.46	3.54
45B	Lymphoma	114.72	107.09	7.63	93.35	6.65
45C	Lymphoma	81.92	81.17	0.75	99.08	0.92
45D	Lymphoma	100.07	96.47	3.60	96.40	3.60
45E	Lymphoma	65.62	56.05	9.57	85.42	14.58
46	Lymphoma	33.17	21.65	11.52	65.27	34.73
47	Lymphoma	18.16	17.93	0.23	98.74	1.26
48	Lymphoma	11.76	8.06	3.70	68.54	31.46
49	Lymphoma	33.86	28.29	5.57	83.56	16.44
50[d]	Osteosarcoma	53.43	16.83	36.60	31.50	68.50
51	Osteosarcoma	14.10	6.88	7.22	48.80	51.20
52[d]	Osteosarcoma	66.47	22.73	43.74	34.19	65.81
53	Osteosarcoma	36.23	32.64	3.59	90.08	9.92
54	Osteosarcoma	60.62	34.35	26.27	56.66	43.34
55[d]	Osteosarcoma	67.46	20.48	46.98	30.36	69.64
56A	Osteosarcoma	117.83	4.82	113.01	4.09	95.91
56B	Osteosarcoma	78.12	27.36	50.76	35.03	64.97
56C	Osteosarcoma	74.28	11.25	63.03	15.15	84.85
57A	Osteosarcoma	49.51	28.83	20.68	58.22	41.78
57B	Osteosarcoma	35.39	0.48	34.91	1.35	98.65
57C	Osteosarcoma	56.11	31.79	24.32	56.66	43.34
58A	Osteosarcoma	100.87	2.32	98.55	2.30	97.70
58B	Osteosarcoma	82.88	11.19	71.69	13.50	86.50
59A	Osteosarcoma	38.31	18.88	19.43	49.27	50.73
59B	Osteosarcoma	61.86	11.11	50.75	17.96	82.04
60A	Osteosarcoma	37.33	2.64	34.69	7.08	92.92
60B	Osteosarcoma	50.39	20.06	30.33	39.81	60.19

TABLE 2

Continued

Patient number[a]	Diagnosis[b]	Tissue area (mm^2)[c]	Tumor area (mm^2)	Stroma area (mm^2)	Percent tumor	Percent stroma
60C	Osteosarcoma	38.56	26.81	11.75	69.53	30.47
60D	Osteosarcoma	32.18	28.14	4.04	87.46	12.54
61A	Osteosarcoma	63.75	35.41	28.34	55.54	44.46
61B	Osteosarcoma	29.38	7.01	22.37	23.86	76.14
62A	Osteosarcoma	14.16	4.75	9.41	33.56	66.44
62B	Osteosarcoma	12.78	2.68	10.10	20.96	79.04
63A	Osteosarcoma	39.24	23.63	15.61	60.21	39.79
63B	Osteosarcoma	52.72	35.42	17.30	67.19	32.81
64	Osteosarcoma	24.00	4.90	19.10	20.42	79.58
65	Osteosarcoma	34.13	15.39	18.74	45.09	54.91
66	Osteosarcoma	39.05	28.88	10.17	73.95	26.05
67	Osteosarcoma	27.32	3.36	23.96	12.31	87.69
68A	Osteosarcoma	35.29	16.65	18.64	47.18	52.82
68B	Osteosarcoma	25.88	6.43	19.45	24.85	75.15
69A	Osteosarcoma	25.39	3.84	21.55	15.12	84.88
69B	Osteosarcoma	4.73	0.85	3.88	17.95	82.05
70A	Osteosarcoma	51.04	1.82	49.22	3.57	96.43
70B	Osteosarcoma	32.70	8.95	23.75	27.38	72.62
70C	Osteosarcoma	74.21	4.00	70.21	5.39	94.61
71A	Osteosarcoma	24.08	2.79	21.29	11.59	88.41
71B	osteosarcoma	33.52	1.85	31.67	5.52	94.48
71C	Osteosarcoma	37.45	4.85	32.60	12.96	87.04
72A	Osteosarcoma	29.08	6.93	22.15	23.84	76.16
72B	Osteosarcoma	23.40	3.55	19.85	15.16	84.84
73A	Osteosarcoma	28.28	4.45	23.83	15.72	84.28
73B	Osteosarcoma	11.21	0.51	10.70	4.55	95.45
73C	Osteosarcoma	9.20	0.37	8.83	3.99	96.01
74A	Osteosarcoma	78.34	31.77	46.57	40.55	59.45
74B	Osteosarcoma	75.72	42.60	33.12	56.26	43.74
74C	Osteosarcoma	147.32	84.72	62.60	57.51	42.49
75A	Osteosarcoma	26.22	15.90	10.32	60.65	39.35
75B	Osteosarcoma	29.66	27.11	2.55	91.41	8.59
75C	Osteosarcoma	25.06	17.58	7.48	70.14	29.86
76A	Osteosarcoma	29.49	26.23	3.26	88.94	11.06
76B	Osteosarcoma	39.96	30.34	9.62	75.93	24.07
77A	Osteosarcoma	10.97	1.46	9.51	13.28	86.73
77B	Osteosarcoma	22.27	3.67	18.60	16.47	83.53
77C	Osteosarcoma	47.39	5.50	41.89	11.60	88.40
78A	Osteosarcoma	65.72	59.74	5.98	90.91	9.09
78B	Osteosarcoma	62.76	49.79	12.97	79.33	20.67
79A	Osteosarcoma	15.48	12.78	2.70	82.55	17.45
79B	Osteosarcoma	23.12	23.05	0.07	99.69	0.31
79C	Osteosarcoma	30.76	27.08	3.68	88.04	11.96
80A	Osteosarcoma	60.44	20.97	39.47	34.70	65.30
80B	Osteosarcoma	26.94	14.97	11.97	55.57	44.43
80C	Osteosarcoma	29.52	20.44	9.08	69.25	30.75
81	Osteosarcoma	138.66	53.57	85.09	38.64	61.36
82	Osteosarcoma	90.71	29.54	61.17	32.57	67.44

TABLE 2

Continued

Patient number[a]	Diagnosis[b]	Tissue area (mm^2)[c]	Tumor area (mm^2)	Stroma area (mm^2)	Percent tumor	Percent stroma
83	Osteosarcoma	27.77	17.54	10.23	63.16	36.84
84	Osteosarcoma	109.09	88.57	20.52	81.19	18.81
85	Osteosarcoma	29.71	11.69	18.02	39.35	60.65
86	Osteosarcoma	47.19	14.23	32.96	30.15	69.85
87	Osteosarcoma	84.90	67.38	17.52	79.37	20.63
88	Osteosarcoma	57.32	2.55	54.77	4.45	95.55
89A	Osteosarcoma	140.01	64.60	75.41	46.14	53.86
89B	Osteosarcoma	85.81	31.89	53.92	37.16	62.84
90	Osteosarcoma	134.21	10.61	123.60	7.91	92.09
91	Osteosarcoma	36.13	31.23	4.90	86.43	13.57
92	Melanoma	27.59	26.78	0.81	97.05	2.95
93	Melanoma	16.56	16.52	0.04	99.79	0.21
94	Melanoma	13.17	13.10	0.07	99.48	0.52
95A	Melanoma	62.83	60.03	2.80	95.55	4.45
95B	Melanoma	54.79	54.48	0.31	99.44	0.56
96	Melanoma	27.39	20.96	6.43	76.53	23.47
97A	Melanoma	12.71	10.56	2.15	83.07	16.93
97B	Melanoma	10.88	9.03	1.85	82.98	17.02
97C	Melanoma	12.43	9.63	2.80	77.46	22.54
98	Melanoma	12.26	6.00	6.26	48.95	51.05
99	Melanoma	17.44	11.01	6.43	63.14	36.86
100A	Melanoma	21.36	20.19	1.17	94.51	5.49
100B	Melanoma	17.10	16.14	0.96	94.39	5.61
100C	Melanoma	16.63	11.23	5.40	67.51	32.49
101A	Melanoma	13.08	8.97	4.11	68.60	31.40
101B	Melanoma	9.72	9.12	0.60	93.84	6.16
102A	Melanoma	46.67	26.74	19.93	57.29	42.71
102B	Melanoma	35.61	2.41	33.20	6.76	93.24
102C	Melanoma	32.95	8.40	24.55	25.50	74.50
103A	Melanoma	6.06	5.01	1.05	82.64	17.36
103B	Melanoma	10.27	7.63	2.64	74.32	25.68
103C	Melanoma	13.55	13.03	0.52	96.14	3.86
104A	Melanoma	68.01	64.65	3.36	95.06	4.94
104B	Melanoma	44.53	39.35	5.18	88.38	11.62
104C	Melanoma	59.99	44.32	15.67	73.88	26.12
105A	Melanoma	4.32	3.80	0.52	87.89	12.11
105B	Melanoma	5.98	4.84	1.14	80.95	19.05
106A	Melanoma	6.07	5.67	0.40	93.39	6.61
106B	Melanoma	33.28	29.42	3.86	88.40	11.60
107A	Melanoma	6.30	3.09	3.21	49.01	50.99
107B	Melanoma	5.82	3.10	2.72	53.28	46.72
107C	Melanoma	13.80	4.66	9.14	33.75	66.25
108A	Melanoma	6.21	4.52	1.69	72.83	27.17
108B	Melanoma	2.79	2.17	0.62	77.80	22.20
108C	Melanoma	2.26	1.40	0.86	61.94	38.06
109	Melanoma	61.11	58.09	3.02	95.05	4.95
110	Melanoma	8.37	5.46	2.91	65.20	34.80
111	Melanoma	207.90	198.59	9.31	95.52	4.48

TABLE 2

Continued

Patient number[a]	Diagnosis[b]	Tissue area (mm^2)[c]	Tumor area (mm^2)	Stroma area (mm^2)	Percent tumor	Percent stroma
112	Melanoma	21.17	6.76	14.41	31.92	68.08
113	Melanoma	68.19	43.72	24.47	64.12	35.88
114	Melanoma	28.32	23.85	4.47	84.21	15.79
115	Melanoma	15.96	12.47	3.49	78.12	21.88
116	Melanoma	22.25	19.03	3.22	85.54	14.46
117	Melanoma	14.81	14.43	0.38	97.40	2.60
118	Melanoma	123.89	105.73	18.16	85.34	14.66
119	Melanoma	21.77	16.99	4.78	78.04	21.96
120	Melanoma	206.21	177.81	28.40	86.23	13.77
121	Melanoma	81.44	71.02	10.42	87.20	12.80
122	Melanoma	70.01	64.36	5.65	91.93	8.07
123	Melanoma	58.93	49.83	9.10	84.56	15.44
124	Melanoma	162.36	124.56	37.80	76.72	23.28
125	Melanoma	39.03	30.84	8.19	79.01	20.99
126	Melanoma	10.16	5.47	4.69	53.84	46.16
127	Melanoma	17.88	15.75	2.13	88.07	11.93
128	Melanoma	21.94	20.02	1.92	91.25	8.75
129	Melanoma	42.76	21.88	20.88	51.18	48.82
130	Melanoma	21.43	4.05	17.38	18.89	81.11
131	Melanoma	65.18	45.85	19.33	70.34	29.66
132A	Melanoma	55.58	51.23	4.35	92.18	7.82
132B	Melanoma	105.18	98.37	6.81	93.52	6.48

mm^2, Square millimeters. [a]Numbers represent individual patients, and letters represent unique tissue sections from single patients. [b]Cancer diagnosis corroborated by a pathologist as part of quality assurance. [c]All values are rounded to the nearest hundredths. Total tissue area rounded to equal tumor + stroma tissue areas. [d]Tissue sections of Patients 50, 52, and 55 had areas of non-neoplastic cartilage included in the section, which were manually annotated and added to the automated stromal area measurements.

DISCUSSION

The techniques developed in this study permitted automated tissue feature area measurements to be obtained from whole-slide image data files of cancer biorepository specimens. By applying unique cancer-specific pattern recognition image analysis algorithms, relevant proportions of neoplastic cells and stromal elements were generated for individual specimens among biorepository collections for multiple tumor types. This demonstrates the feasibility of annotating complex mixtures of cells comprising individual biorepository tissue specimens more precisely than current approaches using visual assessments accomplished by multiple pathologists from multiple institutions over time, which are subjective and can be sources of considerable intra- and inter-observer inconsistency.[8,9]

Systematic evaluation of randomly selected tissue sections from each analyzable biospecimen provided quantifiable tumor area estimates. Area measurements were acquired from diagnostically valid tissue sections used to corroborate the patients' original diagnoses. These analyses not only served for annotating individual banked specimens, but summative frequency distribution analyses of these findings provided an additional perspective about the relative contributions of tumor and stroma that an investigator may expect to encounter across the biorepository collection for each cancer type. Having optimized analyses of relevant cell types and tissue components, data from individual specimens were compared, yielding insight into the untested portions of banked biospecimens from which tissue slices were obtained. For lymphoma and melanoma, median within-patient CVs $\leq$12% and median tumor content areas $\geq$84% imply that there is general consistency within these tumor collections. By contrast, the median portion of tissue area occupied by tumor in the osteosarcoma patients was 39%, with a broad range of tumor tissue content (IQR=44%); these circumstances imply that a number of osteosarcoma biospecimens would likely not meet the caHUB guidelines of 70% tumor, possibly necessitating adjunct approaches for generating neoplastic tissue-relevant biomarkers.[18] Thus, area estimates acquired using single or a few tissue slices/specimen from biorepositories appear to be suitable to gauge features for some

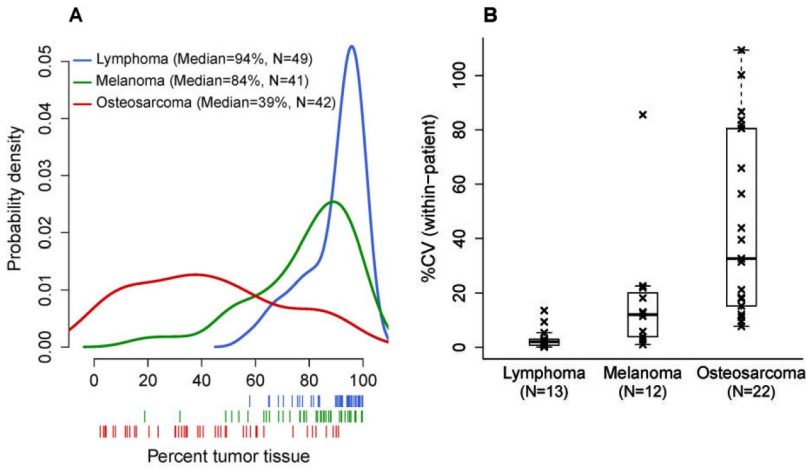

FIGURE 2

Frequency distributions and dispersion analyses of the proportion of cancer biorepository tissue sections occupied by tumor, by cancer type. (*A*) Frequency distribution using kernel density estimation of tumor tissue area percentage for lymphoma (blue line), melanoma (green line), and osteosarcoma (red line) patients. Below the *x*-axis, individual estimates of the tumor tissue percentage are shown as vertical lines. Single tissue sections randomly chosen from all patients with analyzable specimens were used in these analyses (*N*, shown on figure). If more than one specimen was available for a patient, individual tissue sections were chosen randomly from among the multiple specimens. (*B*) Box-whisker plots showing distributions of the within-patient coefficient of variation (%CV) of the tumor tissue percentage estimates in the three cancer types. The line in the middle of the box represents the median, and the lower and upper edges show the first and third quartiles, respectively. The whiskers extend to the extreme values, and the outliers are the CVs exceeding 1.5 × interquartile range (IQR). If patients had more than three specimens available, specimens in excess of three were randomly excluded for statistical evaluation purposes.

tumor types in a collection, similar to the manner that biopsy examination serves as an indication of a patient's disease.[2] However, as cancers contain heterogeneous tumor cells admixed with supporting tissues,[2,5] it is recognized that using tissue section surface area measurements to represent volumetric tissue specimens that exist in biorepositories will not assure definitive comprehension of all specimens' characteristics. Although evaluation of multiple tissue sections from a specimen would likely improve the discernment, histological examination remains a surrogate for the tissue as a whole. The perspective gained by cross-sectional evaluations of biobank specimens can allow for investigators to anticipate the potential variability that could be present in some biobank specimens and to develop additional investigative or quality assurance mechanisms to manage this variability.

These methods, developed using cancers from CCOGC, are directly translatable to human tissue biobank quality-assurance protocols. Spontaneous cancers in pet dogs have similar biology, clinical consequences, and histologic features, including the marked tissue heterogeneity, as human cancer.[10,19,20] As would likely be the case for each biorepository, unique algorithms, ideally trained on locally available biospecimens, would need to be optimized for each human cancer. Additionally, ongoing evaluation of data output from the image analyses by pathologists performing other aspects of quality assurance for the biorepository would remain a necessity.

Limitations exist in applying automated image analyses for biobank pathology quality assurance. Automated image analyses yielded different within-tumor type and within-patient analytic precision, chiefly as a result of inherent phenotypic heterogeneity of different cancers. In general, these algorithms were best for discriminating highly divergent tissues, such as solid sheets of neoplastic lymphocytes and the surrounding stroma. All lymphoma specimens exemplified this phenotype, exhibiting minimal variation in tumor percentage. In contrast, it was more challenging to generate algorithms capable of discriminating composite tissues equally well, such as intermixed neoplastic osteoblasts and osteoid in the case of osteosarcomas, or for tissues with subtle anatomic differences, such as reactive stroma and amelanotic spindloid melanoblasts in some melanomas. Thus, phenotypic heterogeneity among biobank specimens played a role in how suitable a single algorithm served for all lesions within a tumor type. For more heterogeneous histologies, a second algorithm, based on lesions less well-segmented by the first algorithm, was added, permitting image analysis of 76/77 (98.7%) osteosarcomas and 60/70 (85.7%) melanomas. Independently discriminating necrotic neoplasm from overlapping histologic features of fibrosis and osteoid was equally challenging. The difficulties mapping beyond tumor and stroma in this study arose from overlapping spatial-spectral features of tissue classes, which could not be discriminated consistently by image analysis. Additional limitations were encountered as a result of the allowable complexity and size of the montages. Based on the findings reported here, as well as the work of others,[21] further method refinement may be necessary to address inconsistencies in analyzing different

cancers with the same level of precision. It is conceivable that software improvements permitting further enhancement of histological pattern recognition algorithms and use of larger, more representative training sets may help to address these concerns.

Spontaneous cancer biospecimens are composed of a complex mixture of cells and tissue, including tumor cells, stroma, inflammatory cells, adjacent normal tissue, and necrotic tissue.[2,5] As such, macromolecules, including DNA, RNA, and protein, extracted from these tissues, will be derived from and reflect the heterogeneous composition of the tissues. Understanding specimen complexity is pivotal to successful molecular profiling.[1,22] The broad range of tumor percentages among biobank specimens identified in this study for some cancers highlights the value of appreciating specimen complexity. Biorepository annotations provided through our approach will likely impact biomarker discovery and development by helping to define and guide the selection of study-appropriate specimens, thereby reducing wasted resources when conducting downstream studies whose end results are influenced by the characteristics and qualities of the specimens.[1,2,4] The analytic approach developed provides means to delineate some sources of preanalyitic variation as a result of in vivo characteristics of specimens.[23] Image analysis serves as an adjunct to the pathologist's examination but does not eliminate all method subjectivity, which can be introduced during histologic feature selection while developing algorithms.[24] Notwithstanding, applying tumor-specific algorithms uniformly over the course of maintaining a biorepository provides automation, enhances throughput, and has potential to reduce inter- and intra-operator variation compared with a pathologist evaluation alone.

ACKNOWLEDGMENTS

This research was supported by the Intramural Research Program, Center for Cancer Research, NCI (Bethesda, MD, USA). The authors appreciatively acknowledge support and specimens from the Pfizer CCOGC (Rockville, MD, USA). Additional CCOGC sponsors include the Morris Animal Foundation and American Kennel Club Canine Health Foundation. Project logistical support was graciously provided by Christina Mazcko. We thank Jennifer Edwards for scholarly discussions.

REFERENCES

1. Riegman PHJ, Dinjens WNM, Oosterhuis JW. Biobanking for interdisciplinary clinical research. *Pathobiology* 2007;74:239–244.
2. Shevde LA, Riker AI. Current concepts in biobanking: development and implementation of a tissue repository. *Front Biosci (Schol Ed)* 2009;1:188–193.
3. Bevilacqua G, Bosman F, Dassesse T, et al. The role of the pathologist in tissue banking: European Consensus Expert Group Report. *Virchows Arch* 2010;456:449–454.
4. Hallmans G, Vaught JB. Best practices for establishing a biobank. *Methods Mol Biol* 2011;675:241–260.
5. Espina V, Heiby M, Pierobon M, Liotta LA. Laser capture microdissection technology. *Expert Rev Mol Diagn* 2007;7:647–657.
6. Sandusky G, Dumaual C, Cheng L. Human tissues for discovery biomarker pharmaceutical research: the experience of the Indiana University Simon Cancer Center-Lily Research Labs tissue/fluid biobank. *Vet Pathol* 2009;46:2–9.
7. Boudou-Rouquette P, Touibi N, Boelle P-Y, Tiret E, Flejou J-F, Wendum D. Imprint cytology in tumor tissue bank quality control: an efficient method to evaluate tumor necrosis and to detect samples without tumor cells. *Virchows Arch* 2010;456:443–447.
8. Potts SJ, Young GD, Voelker FA. The role and impact of quantitative discovery pathology. *Drug Discov Today* 2010;15:943–950.
9. Brennan DJ, Brändstedt J, Rexhapaj E, et al. Tumour-specific HMG-CoAR is an independent predictor of recurrence free survival in epithelial ovarian cancer. *BMC Cancer* 2010;10:125.
10. Khanna C, Lindblad-Toh K, Vail D. The dog as a cancer model. *Nat Biotechnol* 2006;24:1065–1066.
11. Angeletti C, Harvey NR, Khomitch V, Fischer AH, Levenson RM, Rimm DL. Detection of malignancy in cytology specimens using spectral-spatial analysis. *Lab Invest* 2005;85:1555–1564.
12. Silverman BW. *Density Estimation.* London, UK: Chapman and Hall, 1986.
13. Shapiro SS, Wilk MB. An analysis of variance test for normality (complete samples). *Biometrika* 1965;52:591–611.
14. D'Agostino RB. Transformation to normality of the null distribution of G1. *Biometrika* 1970;57:679–681.
15. Anscombe FJ. Graphs in statistical analysis. *Am Stat* 1973;27:17–21.
16. Tukey JW. *Exploratory Data Analysis.* Reading, MA, USA: Addison-Wesley, 1977.
17. R Development Core Team. R: a language and environment for statistical computing. In *R Foundation for Statistical Computing*. Vienna, Austria, 2010.
18. Liu A. Laser capture microdissection in the tissue biorepository. *J Biomol Tech* 2010;21:120–125.
19. Vail DM, MacEwen EG. Spontaneously occurring tumors of companion animals as models for human cancer. *Cancer Invest* 2002;18:781–792.
20. Paoloni M, Khanna C. Translation of new cancer treatments from pet dogs to humans. *Nat Rev Cancer* 2008;8:147–156.
21. Lloyd MC, Allam-Nandyala P, Purohit CN, Burke N, Coppola D, Bui MM. Using image analysis as a tool for assessment of prognostic and predictive biomarkers for breast cancer: how reliable is it? *J Pathol Inform* 2010;1:29.
22. Botling J, Micke P. Biobanking of fresh frozen tissue from clinical surgical specimens: transport logistics, sample selection and histologic characterization. *Methods Mol Biol.* 2011;275:299–306.
23. Betsou F, Barnes R, Burke T, et al. Human biospecimen research: experimental protocol and quality control tools. *Cancer Epidemiol Biomarker Prev* 2009;18:1017–1025.
24. Tadrous PJ. On the concept of objectivity in digital image analysis in pathology. *Pathology* 2010;42:207–211.