

Debate

Open Access

Tumor classification: molecular analysis meets Aristotle

Jules J Berman*

Address: Cancer Diagnosis Program, National Cancer Institute, Bethesda, USA

Email: Jules J Berman* - bermanj@mail.nih.gov

* Corresponding author

Published: 17 March 2004

BMC Cancer 2004, 4:10

Received: 15 November 2003

Accepted: 17 March 2004

This article is available from: <http://www.biomedcentral.com/1471-2407/4/10>

© 2004 Berman; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

Abstract

Background: Traditionally, tumors have been classified by their morphologic appearances. Unfortunately, tumors with similar histologic features often follow different clinical courses or respond differently to chemotherapy. Limitations in the clinical utility of morphology-based tumor classifications have prompted a search for a new tumor classification based on molecular analysis. Gene expression array data and proteomic data from tumor samples will provide complex data that is unobtainable from morphologic examination alone. The growing question facing cancer researchers is, "How can we successfully integrate the molecular, morphologic and clinical characteristics of human cancer to produce a helpful tumor classification?"

Discussion: Current efforts to classify cancers based on molecular features ignore lessons learned from millennia of experience in biological classification. A tumor classification must include every type of tumor and must provide a unique place for each tumor within the classification. Groups within a classification inherit the properties of their ancestors and impart properties to their descendants. A classification was prepared grouping tumors according to their histogenetic development. The classification is simple (reducing the complexity of information received from the molecular analysis of tumors), comprehensive (providing a place for every tumor of man), and consistent with recent attempts to characterize tumors by cytogenetic and molecular features. The clinical and research value of this historical approach to tumor classification is discussed.

Summary: This manuscript reviews tumor classification and provides a new and comprehensive classification for neoplasia that preserves traditional nomenclature while incorporating information derived from the molecular analysis of tumors. The classification is provided as an open access XML document that can be used by cancer researchers to relate tumor classes with heterogeneous experimental and clinical tumor databases.

Background

Challenge: creating a molecular classification of cancer

In January 1999, the U.S. National Cancer Institute (NCI) issued a challenge to the scientific community "to harness the power of comprehensive molecular analysis technologies to make the classification of tumors vastly more informative. This challenge is intended to lay the groundwork for changing the basis of tumor classification from

morphological to molecular characteristics." [1] Not surprisingly, this has resulted in lively debate over the relative value of morphologic and molecular classifications[2].

What is a tumor classification?

A classification is an organization of everything in a domain by hierarchical groups, according to features generalizable to the members of the groups. Four terms with

distinctly different meanings have been used interchangeably with "classification," leading to considerable confusion among pathologists and cancer researchers. These terms are: identification, discrimination, taxonomy, and ontology [3]. Identification (also known as diagnosing or naming) is the act of placing something into its correct slot within an existing classification. Discrimination is finding features that separate members of a group according to expected variations in group behavior. Examples of discrimination are "grading and staging." Grading and staging involve reporting additional morphologic features (grading) or clinical behavior (staging) that help predict a particular tumor's clinical course or response to therapy. A taxonomy is a complete listing of all the members of a classification. In the case of neoplasia, a taxonomy would be the complete listing of all the different named tumors. An ontology is a rule-based grouping of some portion of a taxonomy. Ontologies support queries and logical inferences pertaining to the [ontologic] group members.

Much of the current work in the molecular classification of tumors is actually discriminant analysis disguised as classification. In a typical gene expression array study, the researcher will look at a group of tumors of a specific type. Cluster analysis of the gene expression array values will help separate the tumors into groups with common expression patterns. Some of these groupings will prove to have a specific biologic feature (e.g. increased tendency to metastasize, higher response to a chemotherapeutic agent, lengthened survival) [4-8]. The groupings seldom qualify as new classes if they merely represent variations in the expected biology of a type of tumor. Variant groups are disqualified as classes if it can be shown that a tumor of a certain type may progress from one variant group to another variant group over time (e.g. slow-growing variant at one stage in development and fast-growing variant at another stage). A key concept in a classification is that the members of one class cannot transform into the members of another class (i.e. a colon carcinoma does not transform into a colon lymphoma).

In the author's opinion, common misuses of the term "classification" form the greatest impediment to progress in the field of cancer genetics. It is impossible to create a molecular classification of tumors based solely on the separation of tumors by variations of molecular markers. Clustering by variation only identifies differences among tumors and is not sufficient to establish a classification. Classification is the process of showing that certain differences reliably distinguish the members of a group from the members of all other groups, and that these differences apply to the group's hierarchical descendants. Therefore, the data that comes from the molecular analysis of tumors can be considered a first-step in the process of tumor classification.

Who actually uses tumor classifications? The author considers himself an example of someone who needs to have a comprehensive tumor classification. As Program Director for Pathology Informatics within the National Cancer Institute's (NCI's) Cancer Diagnosis Program, I am responsible for developing research initiatives that assemble and organize large amounts of pathology data. These efforts need to interoperate with other NCI programs, including the NCI Center for Bioinformatics, which has created tools for linking experimental data, pathology data and clinical trial data. A tumor classification serves as the "key" data structure that links the names of tumors to tumor-related data held in NCI's different clinical and experimental databases. A good classification should help drive-down the complexity of enormous databases and help us discover relationships among different data elements by assembling data under sensible group hierarchies.

The importance of tumor classification

Classifications are important because class properties are shared among the members of a class, and because members of a class inherit the properties of their ancestors. Simply knowing the class of a bacteria can provide a microbiologist with deep insights relating to the expected growth conditions for the organism and the types of antibiotics that may be effective against the organism.

A classification can be thought of as the encapsulation of all knowledge related to a domain. In a modern classification, the elements of the classification (classes and instances) serve as annotation keys and are capable of relating all data to the classification, regardless of the location of the data. Using classed tumor names [from a standard nomenclature], a search of gene expression array databases might locate gene array data specific to the class.

In the last few years, efforts have begun to characterize tumors based on molecular pathways that will serve as targets for new, non-toxic chemotherapeutic agents. There have been early successes with tumors sensitive to the inhibition of tyrosine kinases (gastrointestinal stromal tumors (GIST) and chronic myelogenous leukemia) with Gleevec [9]. Both these tumors derive from non-endodermal/ectodermal embryonic layers, suggesting that molecular pathways (hence targets for chemotherapy) may be class-dependent.

Current status of tumor classification

At present, there is simply no comprehensive modern tumor classification. A practical, though disappointing explanation for this situation is offered by Diamandopoulos. "Since there are almost limitless varieties of tumors, a complete table of classification would require many pages. Any shortened version is not only necessarily

incomplete but also likely to be confusing." [10] This dour perspective may actually represent the modern pathologist's perspective. Current tumor classifications suffer from the following:

1. Classifications are created piecemeal for specific sites or organ systems. Nobody has published a comprehensive classification, although comprehensive taxonomies have been attempted.
2. Classifications are often based on medical disciplines, rather than on any biologic principles (e.g. classification of dermatologic tumors).
3. A given tumor will appear redundantly when subclassifications are merged.
4. No tumor classification has been prepared in a standard format designed to exchange, merge or analyze heterogeneous biological data

The most widely-used authoritative resources are the World Health Organisation classifications, which list the tumors that occur at different body sites [11]. The problem with an organ system approach to classification is that every organ contains organ-specific and organ non-specific cell types. The brain, for instance, contains connective tissue and lymphoid tissue, and therefore is prone to tumors of connective tissue and lymphoid tissue. A listing of tumors that occur in the brain must include: osteocartilaginous tumors, lipoma, fibrous histiocytoma, hemangiopericytoma, rhabdomyosarcoma, melanoma, lymphoma and myeloma, among others. These same tumors will be included again and again in every site-specific classification. Although each term may occur only once in each site-specific classification, the same lesion may occur a virtually limitless number of times when the site classifications are combined into a comprehensive classification of tumors.

Although cancer taxonomies are different from classifications, they usefully provide all the instances of tumors that must be grouped within a classification. Excellent tumor taxonomies are now publicly available at no cost.

Unified Medical Language System (UMLS) Metathesaurus

The UMLS Metathesaurus is a collection of medical terms collected from about 100 different nomenclatures. It is the richest source of term synonymy in existence and has a comprehensive and detailed set of neoplasia-related terms. The UMLS Metathesaurus is available from the U.S. National Library of Medicine at: <http://www.nlm.nih.gov/research/umls/>

Medical Subject Headings (MESH)

MESH is a curated hierarchical listing of medical terms and includes a detailed and comprehensive tumor nomenclature. MESH is provided by the U.S. National Library of Medicine at: <http://www.nlm.nih.gov/mesh/filelist.html>

International Classification of Diseases-Oncology (ICD-O)

The ICD-O is prepared by the World Health Organization. It is in its third version and is available at no cost in the U.S. via state central cancer registries. Non-U.S. facilities should contact their health services for information. Additional information can be obtained at the NCI's Surveillance Epidemiology and End Results (SEER) website at: http://training.seer.cancer.gov/module_icdo3/icdo3_home.html

NCI-Thesaurus

The National Cancer Institute curates a collection of terms related to cancer. The thesaurus can be freely downloaded from: <ftp://ftp1.nci.nih.gov/pub/cacore/EVS/>

Methods

The general rules for classification can be summarized:

1. A classification is a hierarchical grouping, with each group defined by the greatest number of taxa (informative features) that can apply to every instance of the group
2. Every instance must fit into the classification, and every instance and group must have exactly one slot in the classification.
3. Instances and groups are separable from other instances and groups by taxa.
4. Every classification must be constantly tested and restructured (groups and instances) as needed

The classification offered in this article is based on developmental histogenesis and is very similar to classifications described in the mid 1950s [12]. The reasons for this approach to tumor classification are:

1. Organs may have multiple embryologic derivatives (e.g. skin contains tissues of ectodermal, neuroectodermal and mesodermal lineages), but any given cell has only one lineage. This means that a histogenetic classification can assign any tumor to a unique position within the classification.
2. For the most part, tumors have a cell developmental stage. For instance, blastomas are thought to arise from a cell type that precedes organ differentiation. Squamous cell carcinomas of skin are tumors that have features of

cell type that developed from one of the embryonic layers (ectoderm). An embryologic approach permits us to assign a cell type and a developmental stage to tumors.

3. A classification based on developmental histogenesis is relevant to the behavior of tumors. Ectoderm and endoderm-derived tumors metastasize via lymphatics. Mesenchyme-derived tumors tend to metastasize by hematogenous spread.

4. A classification based on developmental histogenesis is consistent with modern molecular analysis of tumors. Mesenchyme-derived tumors tend to be characterized by simple fusion genes. Ectoderm and endoderm-derived tumors tend to be genetically unstable and cannot be characterized by a single genetic abnormality. Primitive blastomas share similar markers regardless of the organ of origin.

Results

The classification schema

The complete classification is available as a supplemental XML file with this article. An abbreviated classification, containing developmental divisions, is shown below. The complete classification schema is available as an annotated XML document with this article [see Additional file 1].

embryonic

primitive

primitive_differentiating

totipotent_or_multipotent_differentiating

limited_differentiating

germ cell

primitive_non_differentiating

non_primitive

endoderm_or_ectoderm

endoderm_or_ectoderm_surface

endoderm_or_ectoderm_endocrine

endoderm_or_ectoderm_parenchymal

odontogenic_epithelium

mesoderm

mesenchyme

connective_tissue

muscle

fibrous_tissue

vascular

adipose_tissue

bone_cartilage

heme_lymphoid

non_mesenchymal_mesoderm

coelomic

coelomic_ductal

coelomic_cavities

coelomic_gonadal

sub_coelomic

sub_coelomic_gonadal

sub_coelomic_endocrine

sub_coelomic_nephric

neuroectoderm_neural_plate

neural_tube

neural_tube_parenchyma

neural_tube_lining

neural_crest

peripheral_nervous_system

neural_crest_endocrine

neural_crest_melanocytic

Explanation of the classification

The classification assigns tumors to the different stages in human development, ignoring embryologic categories that are not associated with tumors and combining

embryologic categories that do not serve to distinguish tumors of a given type.

The first division of the classification divides tumors as embryonic (from which the body develops) or trophoblastic (extra-embryonic) lineage. From the embryonic class come the primitive cells (also call totipotent cells) that precede the differentiated cells. The primitive cells give rise to several subclasses of tumors

- Primitive tumors that remain uncommitted. This would include the Ewing family of tumors, which includes Ewing's tumor peripheral neuroectodermal tumors (PNET), and intra-abdominal desmoplastic small-round-cell tumor [13]. The tumor cells may show certain features of differentiation, but the cells are not equivalent to any specific differentiated cell-type, and seem to occupy a primitive developmental state that precedes the development of embryonic layers.

- Germ cell

- Primitive tumors that differentiate. This group can be divided into primitive tumors with multi-lineage differentiation (as in teratomas) and primitive tumors with restricted differentiation (less than three embryonic layers, such as pancreatoblastoma and hepatoblastoma).

The concepts of germ cells and germinal cells are often confusing to pathologists, who use these terms somewhat differently than embryologists. For pathologists, germ cells are the specialized cells that give rise to ova (in the female) and sperm (in the male). The process of differentiation of ova and sperm is no different than the process of differentiation for any other cell type, and the most frequently occurring tumors in this lineage are dysgerminomas (in females) and seminomas (in males). Germ cells are differentiated cells and should not be classified with totipotent, primordial, uncommitted or primitive cells. The term germinal cells is probably best avoided altogether. It is sometimes used to mean totipotent, sometimes as a synonym for germ cells and sometimes to indicate lineage from one of the early germ layers. All three meanings are unrelated, and the term "germinal" is confusing when applied to tumor classification. In the classification, "germ cell" is intended to mean only one thing: the cell lineage that gives rise to differentiated ova or sperm.

In the classification, the class of endodermal and ectodermal derived tumors are combined. This is done simply because there seems to be no biologic, clinical, morphologic, or molecular differences among the tumors derived from either of these germ layers. This division contains most of the commonly occurring tumors of man. Among

the tumors of the endoderm/ectoderm division, there may be some value in subdividing these by functional cell type. The classification separates the epithelial tumors that arise from ectoderm/endoderm surface epithelium (which would include squamous cell carcinoma of skin or bronchus or esophagus, and adenocarcinoma of colon) from tumors that arise from the ectodermal/endodermal organ epithelium (breast carcinoma, salivary gland carcinoma, pancreatic carcinoma, hepatocellular carcinoma) and tumors arising from ectoderm/endoderm endocrine epithelium (thyroid papillary and follicular carcinoma and pituitary adenoma).

The class of mesodermal tumors is particularly confusing to the non-embryologist because it contains all sarcomas, as well as tumors derived from specialized mesodermal epithelium. In human development, the mesoderm is the embryonic layer separating the ectoderm and the endoderm. It gives rise to all of the connective tissue of the body (i.e. the mesenchyme). The mesenchymal tissues are usually divided into the soft tissues (deriving from muscle, fibrous tissue, and vascular tissue), hard tissue (i.e. bone and cartilage lesions) and hematopoietic (which would include all lymphomas and leukemias regardless of their tissue of origin). In addition to the mesenchyme, the mesoderm is capable of creating lumen or cavities lined by specialized mesodermal epithelium. The coelomic cavities become the pleura, peritoneum, pericardium, and joint spaces, lined by mesothelium and synovium. These give rise to mesothelioma and to synovial sarcoma, two tumors that are morphologically similar, composed of both epithelial cells and spindle cells.

A very specialized coelomic lining cell covers the gonads. These cells are morphologically, clinically and genetically distinct from the other coelomic lining cells and are assigned their own subclass, which includes papillary serous carcinoma of ovary.

The coelom is also capable of forming epithelial-lined ducts. These ducts (such as the paramesonephric duct) give rise to the Fallopian tubes and uterus. All uterine and cervical cancer fall in the class of mesodermal, coelomic-ductal tumors. The mesodermal lineage of these tumors is consistent with the variety of mixed epithelial and non-epithelial tumors arising from the uterus (endometrial carcinoma, carcinosarcoma of endometrium, heterologous mixed mesodermal tumors of uterus).

A specialized mesoderm develops subjacent to coelomic cavities. Specialized sub-coelomic mesoderm tissues give rise to the gonads, the adrenal cortex and to the kidneys. Tumors derived from these specialized mesodermal tissues are assigned their own subclasses. The mesoderm that develops as gonadal stroma gives rise to the sex cord-

stromal tumors of the ovary and testis (e.g. granulosa cell tumor, thecal tumors, Sertoli-Leydig tumors, sex cord tumor with annular tubules). The mesoderm that develops as adrenal cortex gives rise to cortical adenomas and carcinomas. The mesoderm that develops as kidney gives rise to renal cell carcinoma and all of its variants.

Cells deriving from neuroectoderm account for brain tumors (neural tube) and tumors derived from the neural crest (peripheral nervous system tumors, some neuroendocrine tumors and melanocytic tumors). True blastic brain tumors (tumors derived from cells that have not demonstrated neuroectodermal differentiation) would not be classified among tumors deriving from neuroectoderm (e.g. medulloblastoma). Not every tumor with a blastoma suffix is derived from primitive cells. Some blastomas are highly anaplastic versions of tumors that derive from differentiated cells. Glioblastoma is a good example. These tumors are closely related to high grade astrocytomas, and they would be classed as tumors of neural tube parenchyma.

Features of the tumor classification

1. Each tumor occurs only once in the classification
2. The classification is comprehensive (e.g. every tumor of man can be placed somewhere within the classification).
3. The classification is simple. One of the purposes of a classification is to drive down the complexity that exists when the domain taxonomy is large. The entire classification is described by under 40 classifiers.
4. Other tumor classifications divide tumors by medical specialty (e.g. dermatologic neoplasms, hematologic neoplasms, thoracic neoplasms, etc.) This classification is based on biologic principles. The classification uses a feature from developmental biology to capture the most important genomic dichotomy in tumor biology, the separation of tumors with simple and characteristic genetic abnormalities from tumors with genetic instability.
5. The classification has "competence." In the field of informatics, competence is the ability to answer questions related to the instances of a data group.
6. The classification is represented as an XML document.
7. It is easy to add subdivisions to the classification. This is important, as the molecular analysis of tumors is likely to provide new taxa.
8. It is easy to move subdivisions of the classification. Classifications are hypothetical re-creations of reality and must be changed as information is accrued.

9. The classification is easily understood by developmental biologists. Developmental biologists are major participants in post-genomic science and need to have tools to relate basic research with clinical exigencies.

10. The classification is compatible with modern theories of the "stem cell" origin of tumors.

11. The classification does not invalidate existing diagnoses found in pathology reports. The medicolegal importance of this feature cannot be exaggerated. This relieves pathologists from reviewing all their prior cases and re-diagnosing them in conformance with a new classification.

12. The classification is an open access document that can be used or criticized freely by the biomedical community.

Discussion

In general, creating a coherent classification is an intellectually demanding process. Aristotle was the first great classifier. Observing that dolphins have a placenta, he reasoned that dolphins are mammals, not fish. This insight was greeted with almost uniform derision for nearly two thousand years. The fortunes of taxonomists have barely improved in the interim. Gould complains that taxonomy is portrayed as the dullest of all fields [14]. "But classifications are not passive ordering devices in a world objectively divided into obvious categories. Taxonomies are human decisions imposed upon nature – theories about the causes of nature's order [14]. In a recent letter to Nature, Thiele and Yeates comment that research funds go into high profile projects (like the human genome project) but miss classification projects [15]. Classification projects are never-ending because class assignments are tentative and subject to continual testing and improvement [16].

Upper level classification

The highest levels of the classification are the primitive tumors (which include the teratomas and the primitive blastic tumors), tumors of endoderm/ectoderm lineage (containing the overwhelming majority of human cancers), tumors of mesodermal lineage (including all sarcomas) and tumors of neuroectodermal lineage.

The most important value of this classification is the disengagement of tumor type and tumor place of origin. A primitive blastoma may occur in the bone or the brain or the lung, but it is classified along with the other primitive blastomas regardless of location. This permits tumors with similar molecular profiles to be classified according to biological attributes rather than anatomic location (e.g. Ewing's tumor family).

The separation of endodermal/ectodermal from mesodermal tumors is one of the most successful categorizations in tumor biology. The sarcomas tend to have simple cytogenetic and molecular markers (typically translocations leading to gene fusions) [17]. Ectodermal/endodermal tumors tend to have complex cytogenetic abnormalities and genetic instability [18]. The sarcomas behave quite differently from endodermal/ectodermal tumors. Sarcomas metastasize via the blood vessels, and the endodermal/ectodermal tumors metastasize through the lymphatics. The morphologies of the tumor classes are different. Sarcomas tend to have a spindle cell appearance while endodermal/ectodermal cells have an epithelial appearance. It seems very likely that the functional molecular pathways responsible for the malignant phenotype in sarcomas will be different from the pathways followed for endodermal/ectodermal tumors, and that this will result in fundamentally different approaches to finding therapeutic targets against molecular pathways in these tumor classes.

The current classification replaces the morphologic and arbitrary dichotomy of epithelial and non-epithelial neoplasms with the histogenetically definitive concepts of ectodermal/endodermal and non-ectodermal/endodermal neoplasms. The class of ectodermal/endodermal tumors contains most of the neoplasms that are morphologically epithelial. However, a variety of epithelial tumors derive from mesoderm (e.g., renal cell carcinoma, adrenal cortical adenoma and papillary carcinomas of ovary), or neuroectoderm (e.g. melanoma, medullary carcinoma of thyroid). It would seem that these mesodermal and neuroectodermal epithelial tumors may display class behavior independent of their epithelial morphology.

Lower level classification

The classification clarifies many of the genetic and molecular oddities in the field of tumor biology. It distinguishes endocrine tumors based on histogenesis, not by function. Adrenal glands should be thought of as two different glands (medulla and cortex) containing two endocrine cell-types, each with its own embryonic lineage. The medulla derives from neuroectoderm and is sometimes associated with genetic syndromes that involve neuroectoderm-classed tumors. The best example is multiple endocrine adenomatosis type 2a (MEN 2a), characterized by the combined presence of pheochromocytoma and medullary thyroid carcinoma. Both tumors have a neuroectodermal endocrine lineage, the former arising from the adrenal medulla and the latter arising from neural crest C-cells (calcitonin-producing cells) that migrate to the thyroid gland. MEN2 is characterized by *ret* gene mutations [19]. A variety of related genetic syndromes are characterized by pheochromocytomas, medullary thyroid

tumors and other neural crest-derivative tumors and carry the *ret* gene mutation [20,21].

The adrenal cortex is mesodermal in lineage and produces adrenal cortical steroid producing cells, strikingly similar to the steroid cell tumors derived from ovarian mesoderm. The occurrence of adrenal cortical tumors in multiple endocrine adenomatosis type 1 (MEN 1) is somewhat anomalous because all the other endocrine tumors associated with this syndrome are of endoderm endocrine lineage. These include pancreatic islet cell tumors, pituitary tumors and parathyroid tumors. Recent evidence suggests that the adrenal cortical adenomas occur secondarily in response to ACTH hypersecretion and are genetically distinct from the endoderm-derived adenomas seen in MEN1 [22].

Germ cell tumors, in the present classification, are placed adjacent to, but separate from, the teratomas and embryonal carcinomas in a developmental stage prior to the developmental of the embryonic layers (ectoderm, endoderm and mesoderm). This is a departure from classifications of ovarian tumors that include germ cell tumors and ovarian teratomas in the same class. It is the author's opinion that germ cells are different from totipotent embryonic cells. The pure germ cell tumors are seminomas and dysgerminomas. When germ cell tumors are found mixed with teratomas, one can infer a transformation between the different cell types (i.e. germ cells giving rise to totipotent embryonic cells or vice versa).

The ovary, in the current classification, has three anlagen: germ cell, coelomic gonadal and sub_coelomic_gonadal. The appendages of the ovary are derived from coelomic_ducts (paramesonephric or mesonephric). None of the ovary is derived from endoderm or ectoderm. This explains the strangeness and the diversity of tumors arising in, on, or adjacent to the ovary.

Mesotheliomas and synovial sarcomas are placed into the class of tumors with mesodermal, coelomic cavity lineage. This classification emphasizes their similar histogenesis and similar morphology as biphasic epithelioid/spindle tumors. Both tumors have similar histochemical features, producing hyaluronic acid and chondroitin sulfate, the lubricants of coelomic cavities [23]. These two tumors, however, have different cytogenetic features. Synovial sarcomas are characterized by *syt-ssx* fusion transcripts [24], while mesotheliomas have complex cytogenetic abnormalities. It is interesting that synovial sarcomas can arise from many different soft tissue locations, including pericardium and pleura (coelomic cavities) [25-28]. This finding suggests that mesotheliomas and synovial sarcomas are closely related.

Renal tumors are separated from the class of ectodermal/endodermal tumors. The histogenetic anlage for the kidney is metanephric mesoderm. Renal tumors clearly belong in a different class than endodermal/ectodermal tumors. Interestingly, a subset of renal tumors is characterized by fusion gene (PRCC to the TFE3 transcription factor gene). Fusion gene neoplasia is a characteristic absent from tumors of ectodermal/endodermal lineage.

The histogenesis of uterine tumors has always presented a special intellectual challenge [29]. This classification of the uterine tumors offers a radical departure from the classic separation of tumors into epithelial and mesenchymal origins. The uterus, like the kidney, has a purely mesodermal lineage, with no contribution from endoderm or ectoderm (the common lineage for mucosal lining cells). Specifically, the uterus is formed from a duct that forms within the mesoderm (the paramesonephric duct). This duct gives rise to the endometrial epithelium as well as the underlying stroma. Consequently, tumors of endometrial and stromal cells share the same classification (sub-coelomic ductal). Like the kidney, this classification ignores morphologic differences (epithelial versus mesenchymal) and creates a grouping in concordance with the observed mixed epithelial/stromal manifestations of some uterine tumors.

Unresolved issues

The largest class of tumors falls into the ectoderm/endoderm class. This class includes the leading causes of death in man (bronchogenic carcinoma, colon adenocarcinoma, breast carcinoma and prostate carcinoma) and the most frequently occurring (though usually non-lethal) tumors of man (squamous cell carcinoma of skin and basal cell carcinoma of skin). Because so many lesions fall into this one category, it might seem that the classification lacks sufficient complexity.

The human body can be envisioned as a topological donut. It is covered by lining cells, with the donut hole lined by endoderm and the donut outer surface lined by ectoderm. The donut pastry would be the mesoderm. Virtually all exposure to toxic and carcinogenic chemicals is via the surface (ectodermal skin or donut surface) or through our aero-digestive tract (endodermally derived lungs and alimentary tract or donut hole-lining cells). Since ectoderm and endoderm are the cells that are most exposed to carcinogens, it's not surprising that most human cancer falls under the class of ectodermal/endodermal tumors. Likewise, since exposure via the ectoderm/endoderm is a lifelong process, it is not surprising that ectodermal/endodermal tumors tend to increase in incidence with age, occurring disproportionately among the elderly. If further research reveals new taxa that can usefully separate tumors of endodermal and ectodermal

lineage, the current classification would accommodate the change. The combined endodermal/ectodermal class would remain intact because tumors of either lineage have features in common that distinguish them from mesodermal or primitive tumors. Each member of the class of endodermal/ectodermal tumors would be assigned a specific subclass.

How do we know if the classification is correct? The correctness of a classification is determined by adding new feature information that characterize classes and instances. When newly added group features extend beyond the class or fail to extend to descendant classes or fail to extend to all the members of a class, the classification needs to be modified. Testing the classification is a never-ending but worthwhile process.

Mayr equates classifications with macrotaxonomies [3]. The taxonomy is the list of objects included in the macrotaxonomy. The macro-taxonomy is the scaffold into which the members of the taxonomy might fit. Provided with this article is a "filled-in" classification, using approximately 55,000 neoplasia terms extracted from the NCI-Thesaurus [see Additional file 2].

Summary

A tumor classification is proposed that groups each tumor according to embryonic lineage. This classification is prepared as an XML file designed to accommodate additional attributes (e.g. genomic, proteomic, clinical information). The classification is comprehensive (can include all neoplastic entities) and parsimonious (each entity has one lineage). This classification blends traditional concepts of tumor nomenclature with post-genomic concepts of neoplastic development. The classification structure and the classification with taxonomic annotation (approximately 55,000 terms) are available as supplemental files with this article.

Competing interests

None declared.

Authors' contributions

This work represents the opinions of the author and does not represent the policy of the NIH or of any other U.S. Federal Agency.

Additional material

[Additional File 1](#)

Neoplasia classification structure (XML version). Neoclass.xml is a pure XML file that can be viewed in current web browsers. It shows the bare classification scheme, with a few sample annotations. The file must have a .xml filename suffix before it can be opened and viewed on a web browser. If this suffix is lost during download, the reader should simply rename the file neoclass.xml to provide the .xml suffix.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2407-4-10-S1.xml>]

Additional File 2

Neoplasia classification with taxonomy (XML version). Neocxml.gz is a compressed (gzipped) pure XML file. If the filename is changed during download, it should be renamed neocxml.gz so that the .gz suffix can be recognized by unzip utilities. Unzip the file (recommended utility: gunzip.exe). Once unzipped, the file is 4 Mbytes in length. The expanded file should be renamed neocl.xml, to provide an XML suffix recognizable to web browsers. It can be opened on current versions of popular web browsers, but because it is a very large file, it may require substantial memory to view. The file contains approximately 55,000 coded neoplasia terms, all assigned to the classification structure.

Click here for GZ version of file

Click here for ZIP version of file

[<http://www.biomedcentral.com/content/supplementary/1471-2407-4-10-S2.gz>]

Acknowledgements

This work was conducted at the NIH as part of the author's customary work activities, and no specific financial support was received for this work.

References

1. Director's challenge: toward a molecular classification of tumors [<http://grants.nih.gov/grants/guide/rfa-files/RFA-CA-98-027.html>]
2. Rosai J: **The continuing role of morphology in the molecular age.** *Mod Pathol* 2001, **14**:258-260.
3. Mayr E: *The growth of biological thought: diversity, evolution and inheritance* Cambridge: Belknap Press; 1982.
4. Eisen MB, Spellman PT, Brown PO, Botstein D: **Cluster analysis and display of genome-wide expression patterns.** *Proc Natl Acad Sci USA* 1998, **95**:14863-14868.
5. Rosenwald A, Wright G, Chan WC, Connors JM, Campo E, Fisher RI, Gascoyne RD, Muller-Hermelink HK, Smeland EB, Giltman JM, Hurt EM, Zhao H, Averett L, Yang L, Wilson WH, Jaffe ES, Simon R, Klausner RD, Powell J, Duffey PL, Longo DL, Greiner TC, Weisenburger DD, Sanger WG, Dave BJ, Lynch JC, Vose J, Armitage JO, Montserrat E, Lopez-Guillermo A, Grogan TM, Miller TP, LeBlanc M, Ott G, Kvaloy S, Delabie J, Holte H, Krajci P, Stokke T, Staudt LM, Lymphoma/Leukemia Molecular Profiling Project: **The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma.** *N Engl J Med* 2002, **346**:1937-1947.
6. Ramaswamy S, Tamayo P, Rifkin R, Mukherjee S, Yeang C, Angelo M, Ladd C, Reich M, Latulippe E, Mesirov JP, Poggio T, Gerald W, Loda M, Lander ES, Golub TR: **Multiclass cancer diagnosis using tumor gene expression signatures.** *Proc Natl Acad Sci USA* 2001, **98**:15149-15154.
7. Weyers W, Euler M, Diaz-Cascajo C, Schill W, Bonczkowitz M: **Classification of cutaneous malignant melanoma: A reassessment of histopathologic criteria for the distinction of different types.** *Cancer* 1999, **86**:288-299.
8. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES: **Molecular classification of cancer: class discovery and class prediction by gene expression monitoring.** *Science* 1999, **286**:531-537.
9. Capdeville R, Silberman S: **Imatinib: A targeted clinical drug development.** *Semin Hematol* 2003, **40**:15-20.
10. Diamandopoulos GT, Meissner WA: **Neoplasia.** In *Anderson's Pathology* Edited by: Kissane JM. St. Louis: Mosby; 1985:518-520.
11. Kleihues P, Burger PC, Scheithauer BW: **The new WHO classification of brain tumours.** *Brain Pathology* 1993, **3**:255-68.
12. Willis RA: **Borderland of Embryology and Pathology.** London: Butterworth 1958.
13. Cohn SL: **Diagnosis and classification of the small round-cell tumors of childhood.** *Am J Pathol* 1999, **155**:11-15.
14. Gould SJ: *Full house: The spread of excellence from Plato to Darwin* New York: Harmony; 1996:38-42.
15. Thiele K, Yeates D: **Tension arises from duality at the heart of taxonomy: Names must both represent a volatile hypothesis and provide a key to lasting information.** *Nature* 2002, **419**:337. inclusive
16. Logsdon JM, Faguy DM: **Evolutionary genomics: Thermotoga heats up lateral gene transfer.** *Current Biology* 1999, **9**:R747-R751.
17. Martine Peter M, Gilbert E, Delattre O: **A Multiplex Real-Time PCR Assay for the Detection of Gene Fusions Observed in Solid Tumors.** *Laboratory Investigation* 2001, **81**:905-912.
18. Hittelman WN: **Genetic instability in epithelial tissues at risk for cancer.** *Ann N Y Acad Sci* 2001, **952**:1-12.
19. Bongarzone I, Vignano E, Alberti L, Borrello MG, Pasini B, Greco A, Mondellini P, Smith DP, Ponder BAJ, Romeo G, Pierotti MA: **Full activation of MEN2B mutant RET by an additional MEN2A mutation or by ligand GDNF stimulation.** *Oncogene* 1998, **16**:2295-2301.
20. Jensen JC, Choyke PL, Rosenfeld M, Pass HI, Keiser H, White B, Travis W, Linehan WM: **A report of familial carotid body tumors and multiple extra-adrenal pheochromocytomas.** *J Urol* 1991, **145**:1040-1042.
21. DeAngelis LM, Kelleher MB, Post KD, Fetell MR: **Multiple paragangliomas in neurofibromatosis: a new neuroendocrine neoplasia.** *Neurology* 1987, **37**:129-133.
22. Skogseid B, Larsson C, Lindgren PG, Kvanta E, Rastad J, Theodorsson E, Wide L, Wilander E, Oberg K: **Clinical and genetic features of adrenocortical lesions in multiple endocrine neoplasia type 1.** *J Clin Endocrinol Metab* 1992, **75**:76-81.
23. Nakamura T, Nakata K, Hata S, Ono K, Katsuyama T: **Histochemical characterization of mucosubstances in synovial sarcoma.** *Am J Surg Pathol* 1984, **8**:429-434.
24. Nishio J, Iwasaki H, Ishiguro M, Ohjimi Y, Isayama T, Naito M, Kikuchi M: **Identification of syt-ssx fusion transcripts in both epithelial and spindle cell components of biphasic synovial sarcoma in small tissue samples isolated by membrane-based laser microdissection.** *Virchows Arch* 2001, **439**:152-157.
25. Anand AK, Khanna A, Sinha SK, Mukherjee U, Walia JS, Singh AN: **Pericardial synovial sarcoma.** *Clin Oncol* 2003, **15**:186-188.
26. Peoch M, Le Marchandour F, Bost F, Pasquier D, Roux JJ, Pinel N, Leroux D, Pasquier B: **Primary synovial sarcoma of the mediastinum. A case report with immunohistochemistry, ultrastructural and cytogenetic study.** *Ann Pathol* 1995, **15**:203-206.
27. Sidhar SK, Clark J, Gill S, Hamoudi R, Crew AJ, Gwilliam R, Ross M, Linehan WM, Birdsall S, Shipley J, Cooper CS: **The t(X;1)(p11.2;q21.2) translocation in papillary renal cell carcinoma fuses a novel gene PRCC to the TFE3 transcription factor gene.** *Human Molecular Genetics* 1996, **5**:1333-1338.
28. Essary LR, Vargas SO, Fletcher CD: **Primary pleuropulmonary synovial sarcoma: reappraisal of a recently described anatomic subset.** *Cancer* 2002, **94**:459-469.
29. Aubry MC, Bridge JA, Wickert R, Tazelaar HD: **Primary monophasic synovial sarcoma of the pleura: five cases confirmed by the presence of SYT-SSX fusion transcript.** *Am J Surg Pathol* 2001, **25**:776-781.
30. Seidman JD, Chauhan S: **Evaluation of the relationship between adenocarcinoma and carcinosarcoma and a hypothesis of the histogenesis of uterine sarcomas.** *Int J Gynecol Pathol* 2003, **22**:75-82.

Pre-publication history

The pre-publication history for this paper can be accessed here:

<http://www.biomedcentral.com/1471-2407/4/10/prepub>