

## Assessing interobserver variability and accuracy in the histological diagnosis and classification of cutaneous neurofibromas

Nicolas Ortonne, Steven L. Carroll, Fausto J. Rodriguez, Douglas C. Miller, Rosalynn M. Nazarian, Jaishri O. Blakeley, Zachary B. Madaj,<sup>®</sup> Sharad K. Verma, and Anat Stemmer-Rachamimov

*Department of Pathology, Henri Mondor Hospital Paris Est Creteil, France (N.O.); Department of Pathology and Laboratory Medicine, Medical University of South Carolina, Charleston, South Carolina (S.L.C.); Department of Pathology, Johns Hopkins School of Medicine, Baltimore, Maryland (F.J.R.); Department of Pathology and Anatomical Sciences, University of Missouri School of Medicine, Columbia, Missouri (D.C.M.); Department of Pathology, Dermatopathology Unit, Massachusetts General Hospital, Boston, Massachusetts (R.M.N.); Department of Neurology, Johns Hopkins School of Medicine, Baltimore, Maryland (J.O.B., S.K.V.); Bioinformatics and Biostatistics Core, Van Andel Research Institute, Grand Rapids, Michigan (Z.B.M.); Department of Pathology, Neuropathology Division, Massachusetts General Hospital, Boston, Massachusetts (A.S.-R.)*

Corresponding Author: Anat Stemmer-Rachamimov, MD, Warren Building, Rm 333A, Massachusetts General Hospital, 55 Fruit Street, Boston, MA 02114 ([astemmerrachamimov@partners.org](mailto:astemmerrachamimov@partners.org)).

### Abstract

**Background.** Cutaneous neurofibromas (cNFs) are the most common tumors in people with neurofibromatosis type 1 (NF1) and are associated with reduced quality of life. There is currently no widely accepted standardized language for describing cNFs clinically or histopathologically. The objective of this study was to evaluate interobserver agreement across pathologists in describing and reporting of neurofibromas involving the skin.

**Methods.** Twenty-eight (H&E)-stained slides of cNF were scanned using an Aperio XT scanner. The digital images were reviewed by 6 pathologists, who entered free text of up to a 200 word description for each case into a REDcap database. Responses were analyzed for the most commonly used terms based on frequency, as well as agreement (reported as concordance) between reviewers.

**Results.** A set of the terms most commonly used by pathologists for the histological classification of cNF along with areas of agreement and disagreement have been identified. The study shows that there was strong agreement across reviewers that not all neurofibromas involving the skin are cutaneous neurofibromas and regarding the presence or absence of atypical features and heterologous elements. Areas of less concordance were identified and include cNF subtypes, definition of extension and pattern of growth, as well as the distinction of a cNF from a plexiform without an intraneural component involving skin.

**Conclusions.** This work is the first step towards development of a robust classification system and devising “gold standard” histopathologic diagnostic criteria for cutaneous neurofibromas.

### Key Points

There was:

- Concordance regarding distinction between a cNF and a plexiform neurofibroma with intraneural component.
- Concordance regarding absence or presence of atypical features in cNF.
- Low concordance regarding cNF subtypes, extent, and pattern of growth.

## Importance of the Study

There is currently no commonly accepted classification scheme for describing cNF. Having such a scheme is essential for physicians in order to have informed discussions with drug developers and regulators and to develop interventional therapies (for which none currently exist) for this

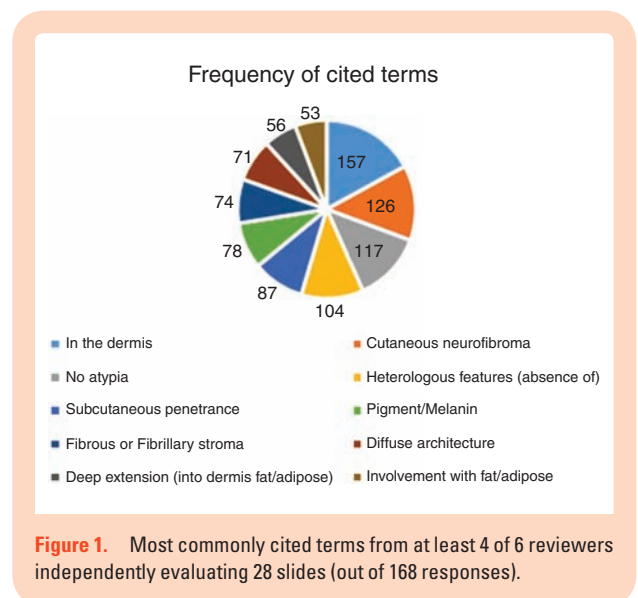
prevalent disease. Gaining a better understanding of how pathologists of different subspecialties view cNFs, and identifying areas of agreement and disagreement amongst their features, represents an important first step towards developing a commonly accepted classification scheme.

Neurofibromatosis type 1 (NF1) is an autosomal dominant genetic disease affecting 1 of 2500 persons<sup>1</sup> characterized by cutaneous, skeletal, neurological, and neoplastic manifestations. In adults, cutaneous neurofibromas (cNFs) are the most common manifestation of the disease, affecting more than 99% of patients.<sup>2-8</sup> Cutaneous neurofibromas increase in number and size with age across the patients' lifespan, most often first appearing between late school age and in the second decade.<sup>2,4,9,10</sup> Although benign, these tumors are a major source of emotional and social distress<sup>2,11-13</sup> and may also cause intermittent but chronic physical symptoms such as pain and itching.<sup>14</sup> Therefore, although benign, cutaneous neurofibromas are highly damaging to NF1 patients via their disfigurement, pain, and itching.

There is no common pathologically based terminology for neurofibromas involving skin, as different terminology is used by various subspecialties (neuropathology, dermatopathology, and soft tissue pathology) and across institutions.<sup>15</sup> Several efforts to define the various forms of neurofibromas have been proposed by neurologists, clinical geneticists, dermatologists, pathologists, and basic researchers, and these have resulted in a number of classification systems from various pathology organizations. For example, guidelines proposed by the College of American Pathologists may be used by general pathologists, whereas neuropathologists rely upon the 2016 WHO classification of tumors of the central nervous system<sup>16</sup> (revised from the 4th edition published in 2007<sup>17</sup>). Dermatopathologists may use recommendations from the American Society of Dermatopathology,<sup>18,19</sup> and soft tissue pathologists the WHO classification guidelines for soft tissue and bone tumors.<sup>20</sup> In some cases, one lesion is described with multiple terms (e.g., nodular, discrete, and localized cutaneous neurofibromas), whereas in other cases the same term may be used to describe two different lesions (e.g., nodular may be intra- or extra-neural). So, despite these valuable guidelines, some of the schemes are incompatible with one another and have hindered rather than improved our understanding of the disease and the development of effective therapies. A summary of the current classifications for cutaneous neurofibromas was recently published.<sup>15</sup> As a first step in developing a classification system for cNFs, the present study assesses the interobserver variability and accuracy of reporting of neurofibromas involving skin by a panel of pathologists of different subspecialties. From this preliminary step, we will be positioned to develop standardized terminology for describing all neurofibromas clinically and on histopathologic examination.

## Methods

The study was approved by the appropriate institutional review boards. Deidentified histopathology pictures of neurofibromas involving the skin from patients diagnosed with NF1 were obtained from the French national referral center (NF1 Ile de France). Samples were biopsied or resected between 01 April 2014 and 31 April 2017 and analyzed in the Department of Pathology of Henri Mondor hospital. Tumors were removed by surgery and/or laser. Tumors with morphological artifacts hampering interpretation were excluded from this study. All samples were paraffin-embedded sections stained with hematoxylin and eosin. Samples analyzed ( $N = 28$ ) were from 12 women and 5 men (age range of 15–62, mean age 44) at the time of biopsy or resection. The lesions ranged from 0.2 to 3.0 cm (mean size 1.2 cm and median size 0.9 cm, measured by soft ruler). The samples were resected from skin from the scalp, fingers, shoulder, arm, cheek, foot, ankle, and back and processed routinely for histopathology. The hematoxylin and eosin (H&E)-stained slides were scanned using an Aperio XT scanner. Digital slides were downloaded using the native image format, Aperio SVS files and analyzed using the CaloPix viewer (TRIBVN-HEALTHCARE, Chatillon, France), and digital images were shared with the reviewers through the internet. Individual reviews were conducted by 4 neuropathologists and 2 dermatopathologists, who evaluated



**Table 1.** Results from independent reviews for frequently cited terms and agreement across reviewers

Term	No. of times cited	Results	Level of concordance
In the dermis	157	Term cited in all slides with $\geq 67\%$ agreement for 27/28 slides (96%) <ul style="list-style-type: none"> <li>•100% agreement for 14 slides</li> <li>•83% agreement for 11 slides</li> <li>•67% agreement for 2 slides</li> <li>•50% agreement for 1 slide</li> </ul>	Strong
Cutaneous neurofibroma	126	Term cited in 26/28 slides (93%) with $\geq 67\%$ agreement for 22/26 slides (85%) <ul style="list-style-type: none"> <li>•83% agreement for 10 slides</li> <li>•67% agreement for 12 slides</li> <li>•50% agreement for 2 slides</li> <li>•33% agreement for 2 slides</li> <li>•17% agreement for 2 slides</li> </ul>	Strong
Atypia or atypical (absence of)	117	Term cited in 27/28 slides (96%) with $\geq 67\%$ agreement for 27/28 slides (96%) <ul style="list-style-type: none"> <li>•83% agreement for 8 slides</li> <li>•67% agreement for 19 slides</li> <li>•50% agreement for 1 slide</li> </ul>	Strong
Heterologous features (absence of)	104	Term cited in all slides, with 67% agreement for 22/28 slides (79%). <ul style="list-style-type: none"> <li>•67% agreement for 22 slides</li> <li>•50% agreement for 6 slides</li> </ul>	Strong
Subcutaneous penetrance	87	Term cited in all slides, with $\geq 67\%$ agreement for 17/28 slides (61%). <ul style="list-style-type: none"> <li>•100% agreement for 8 slides</li> <li>•83% agreement for 4 slides</li> <li>•67% agreement for 5 slides</li> <li>•50% agreement for 3 slides</li> <li>•33% agreement for 5 slides</li> <li>•17% agreement for 3 slides</li> </ul>	Strong
Pigment/melanin	78	Term cited in all slides, with 67% agreement for 4/28 (14%) and $\leq 50\%$ for 24/28 slides (86%). <ul style="list-style-type: none"> <li>•67% agreement for 4 slides</li> <li>•50% agreement for 14 slides</li> <li>•33% agreement for 10 slides</li> </ul>	Mixed
Fibrous or fibrillary stroma	74	Term cited in all slides with agreement limited up to 50% for all slides. <ul style="list-style-type: none"> <li>•50% agreement for 20 slides</li> <li>•33% agreement for 6 slides</li> <li>•17% agreement for 2 slides</li> </ul>	Weak
Diffuse architecture	71	Term cited in all slides with 67% agreement for 5/28 slides (18%) and $\leq 50\%$ agreement for 23/28 slides (82%). <ul style="list-style-type: none"> <li>•67% agreement for 5 slides</li> <li>•50% agreement for 9 slides</li> <li>•34% agreement for 10 slides</li> <li>•17% agreement for 4 slides</li> </ul>	Mixed
Deep extension (into dermis, fat/adipose)	56	Term cited in 24/28 slides (86%) with 67% agreement for 4/24 slides (17%) and $\leq 50\%$ agreement for 15/24 slides (83%) <ul style="list-style-type: none"> <li>•67% agreement for 4 slides</li> <li>•50% agreement for 5 slides</li> <li>•34% agreement for 10 slides</li> <li>•17% agreement for 5 slides</li> </ul>	Mixed
Involvement with fat/adipose	53	Term cited in 22/28 of slides (79%) with 67% agreement for 6/22 slides (28%) and $\leq 50\%$ agreement for 16/22 slides (72%) <ul style="list-style-type: none"> <li>•67% agreement for 6 slides</li> <li>•50% agreement for 5 slides</li> <li>•34% agreement for 7 slides</li> <li>•17% agreement for 4 slides</li> </ul>	Mixed

each image and entered free text of up to a 200 word description for each case into a REDcap database.<sup>21</sup> A total of 168 reviews were received for the 28 slides assessed by the

6 reviewers. Concordance by reviewers was evaluated as % agreement, defined as 100% = 6 of 6 reviewers agree, 83% = 5 of 6 reviewers agree, 67% = 4 of 6 reviewers agree, 50% = 3

of 6 reviewers agree, 33% = 2 of 6 reviewers agree, and 17% = comment made by one reviewer only. Concordance was considered “strong” when >50% of reviewers (i.e., at least 4 reviewers) were in agreement for >50% of the slides evaluated. Responses were analyzed for the most commonly used terms based on frequency, as well as agreement (reported as concordance) between reviewers. Concordance was considered “weak” when agreement from a majority of reviewers (at least 4 of 6) was not ever observed, and 3 or less of 6 reviewers ( $\leq 50\%$ ) were in agreement for a majority of the slides. All other scenarios were considered to have “mixed” concordance.

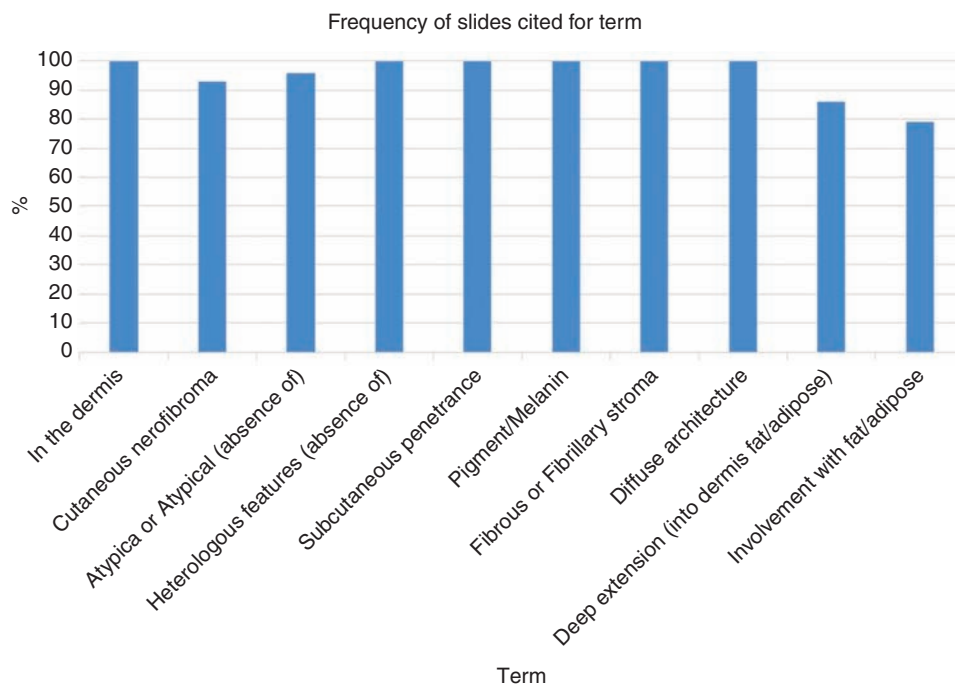
## Results

The 10 terms most commonly cited from at least 4 of 6 reviewers were “in the dermis,” “cutaneous neurofibroma,” “absence of atypia or atypical features,” “absence of heterologous features,” “subcutaneous penetrance,” “pigment/melanin,” “fibrous or fibrillary stroma,” “diffuse architecture,” “deep extension (into dermis, fat/adipose),” and “involvement with fat/adipose” (Figure 1; Table 1). Moreover, these 10 terms were cited in almost all the slides evaluated (Figure 2). Concordance was however variable across reviewers with respect to the slides where these features appeared (Figure 3). Concordance was considered “strong” for the terms “involvement in the dermis,” “cutaneous neurofibroma,” “absence of atypia or atypical features,” “absence of heterologous features,” and “subcutaneous penetrance,” where the majority of reviewers (at least 4 of 6 reviewers) were in agreement for >50% of the slides reviewed. In

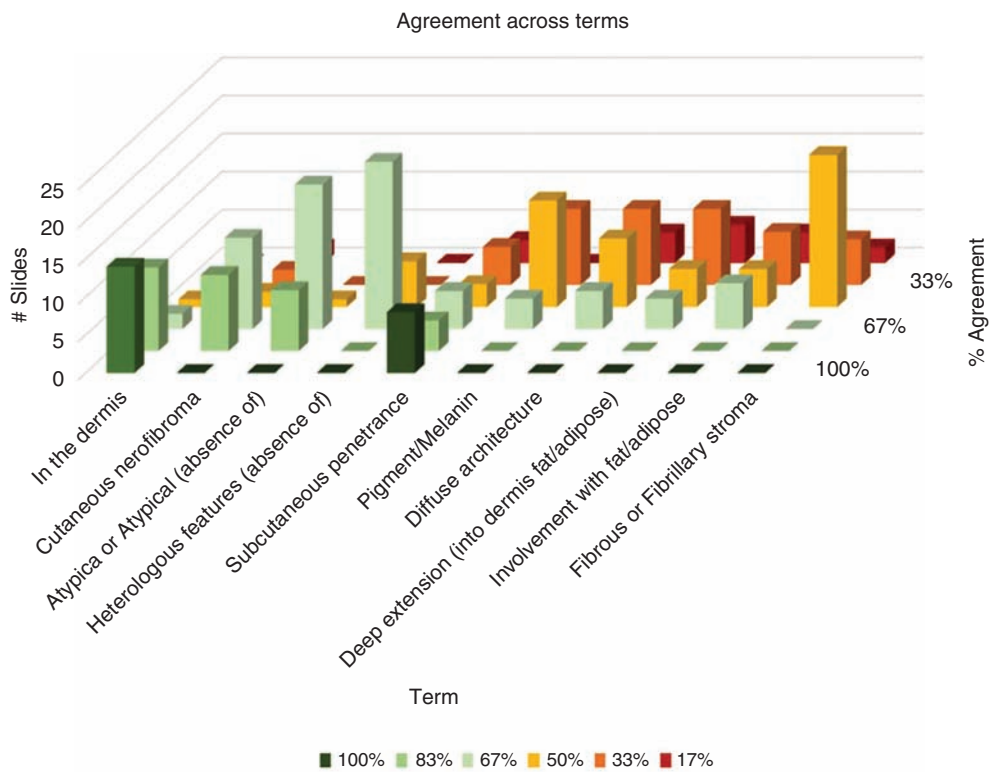
contrast, concordance was considered “weak” for the terms “fibrous or fibrillary stroma” where agreement from the majority of reviewers was not ever observed, and  $\leq 50\%$  of reviewers were in agreement for a majority of the slides. The level of concordance was observed as being “mixed” for the terms, “pigment/melanin,” “diffuse architecture,” “deep extension (into dermis, fat/adipose),” and “involvement with fat/adipose.” The subtype of cNF diagnosed (flat/sessile/globular/pedunculated) was cited 48 times in up to 22 images. However, the level of concordance for diagnoses was low, irrespective of the subtype (Figures 4 and 5; Table 2). For two cases diagnosed as not being cNF (a plexiform neurofibroma with intraneural component extending to the skin), there was strong agreement from reviewers (agreement from at least 4 of the 6 reviewers; Figure 5). Overall, for the diagnosis of cNF, a concordance rate of at least 67% (i.e., agreement from at least 4 reviewers) was observed for 79% of the cases evaluated (22 of 26). For the remaining cases (4 of 26; 21%), the rate of concordance was  $\leq 50\%$ .

## Discussion

cNFs are the most common tumor in people with NF1.<sup>2-8</sup> There is currently no widely accepted standardized language for describing cNFs clinically or histopathologically.<sup>16</sup> The objective of this study was to gain a better understanding of the terms that pathologists commonly use to describe neurofibromas that involve the skin and to evaluate interobserver agreement across experienced pathologists. From this preliminary step, we will be positioned towards developing standardized language for



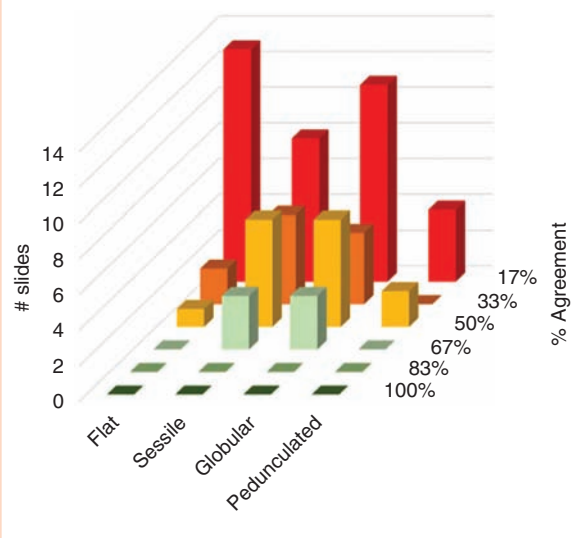
**Figure 2.** Most commonly cited terms appear in almost all the slides reviewed.



**Figure 3.** Levels of agreement vary across the most commonly cited terms.

describing all neurofibromas clinically and histopathologically, which will assist in more accurate clinical diagnosis and in the development of cNF-specific therapeutics. From this study evaluating neurofibromas in the skin, a set of terms most commonly used by pathologists for the histological classification of cNF along with areas of agreement and disagreement have been identified. The study shows that there was strong agreement across reviewers that not all neurofibromas involving the skin are cutaneous neurofibromas. Pathologists were in agreement distinguishing between cNF and plexiform neurofibromas with intraneural component involving the skin and there was also concordance regarding the presence or absence of atypical features and heterologous elements. Areas of less concordance were identified and include cNF subtypes, definition of extension and pattern of growth, as well as the distinction of a cNF from a plexiform involving the skin when an intraneural component is absent.

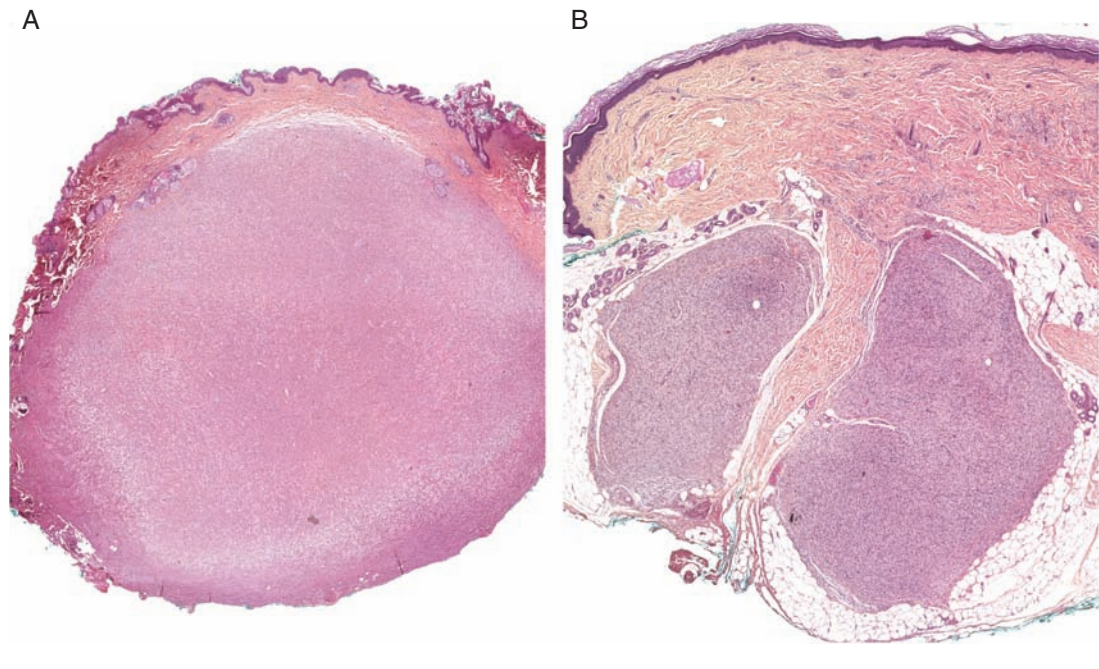
The lack of concordance observed for cNF diagnoses with respect to subtype, based on histological analysis, was highly notable. This may reflect the absence of a universal, accepted classification scheme. Moreover, some of the terms used may have different meanings; for example, the term “diffuse” or the term “deep” may need to be better defined. This study may also point to the need for providing a list of features that should be listed when histologically evaluating a neurofibroma, with possible inclusion of a small panel of immunostains that can aid the pathologist in describing the cellular and stromal components. It is noted that this is a limited study where only 28



**Figure 4.** Less agreement across reviewers with respect to classifications.

slides were evaluated, and so evaluation of a much larger sample set could possibly result in greater concordance. However, the advantage of this set is that the slides were prepared in an identical manner and digitized enabling the analysis of the same images by the pathologists.





**Figure 5.** Histological features of neurofibromas in the skin; H&E stain. Example of a lesion with low concordance regarding histological subtype (A) and a lesion with high concordance (plexiform involving skin; B).

**Table 2.** cNF classification with respect to reviewer agreement

cNF classification	No. of images cited	No. of times classification cited by reviewers	Agreement levels across reviewers
Flat	15	20	50% agreement for 1 case (7%) 33% agreement for 2 cases (14%) No agreement across 12 cases (79%)
Sessile	22	48	67% agreement for 3 cases (14%) 50% agreement for 6 cases (27%) 33% agreement for 5 cases (23%) No agreement for 8 cases (36%)
Globular	19	39	67% agreement for 3 cases (16%) 50% agreement 6 cases (31%) 33% agreement for 4 cases (21%) No agreement across 11 cases (58%)
Pedunculated	6	10	50% agreement for 2 cases (33%) No agreement for 4 cases (67%)

## Conclusion

A “baseline level” of understanding of how pathologists with different subspecialty expertise view cNFs has been obtained. Although there was agreement in the distinction between cNF and plexiform NF (with intraneural component) involving skin, and presence or absence of atypical features, there was less concordance with respect to defining cNF patterns of growth, composition, extent and histological

subtypes, as well as the distinction of a cNF from a plexiform without an intraneural component involving skin. Recognition of the differences in thought conferred by the reviewing pathologists represents an important first step towards the broader goal of accurately describing each tumor type by its clinical classification (location and behavior), clinical appearance (adjective), and molecular and histological features. This work is the first step towards development of a robust classification system and devising “gold standard” histopathologic diagnostic criteria for

cutaneous neurofibromas. Doing so will allow for consistent dialogue pertaining to the correlation of histological features with clinical and molecular data, eventual stratification of neurofibromas, comparisons, and trials with similar animal models, supporting development of tumor specific therapies.

## Keywords

cutaneous neurofibroma | interobserver variability | NF1 | pathology.

## Funding

This work was supported by an agreement from The Johns Hopkins University School of Medicine and the Neurofibromatosis Therapeutic Acceleration Program (NTAP). Its contents are solely the responsibilities of the authors and do not necessarily represent the official views of The Johns Hopkins University School of Medicine.

## Acknowledgments

The authors are grateful to Dr. Pierre Wolkenstein for many helpful discussions. The materials presented were shared at the 2019 Children's Tumor Foundation NF conference.

**Conflict of interest statement.** None declared.

## Authorship Statement

A.S.-R, J.O.B., and N.O. conceived of the experimental design. N.O. assembled the slides for review. N.O, S.L.C., F.J.R., D.C.M., R.M.N., and A.S.-R. conducted the experimental work, reviewed the analysis and interpretation of the data, and contributed to the writing of the manuscript. J.O.B. contributed the writing of the manuscript. S.K.V. compiled the data, conducted analysis and interpretation of the data, and wrote the manuscript.

## References

1. Anderson JL, Gutmann DH. Neurofibromatosis type 1. *Handb Clin Neurol*. 2015;132:75–86.
2. Duong TA, Bastuji-Garin S, Valeyrie-Allanore L, Sbidian E, Ferkal S, Wolkenstein P. Evolving pattern with age of cutaneous signs in neurofibromatosis type 1: a cross-sectional study of 728 patients. *Dermatology*. 2011;222(3):269–273.
3. Wolkenstein P, Zeller J, Revuz J, Ecosse E, Lepège A. Quality-of-life impairment in neurofibromatosis type 1: a cross-sectional study of 128 cases. *Arch Dermatol*. 2001;137(11):1421–1425.
4. Friedman JM. Epidemiology of neurofibromatosis type 1. *Am J Med Genet*. 1999;89(1):1–6.
5. Lammert M, Friedman JM, Kluwe L, Mautner VF. Prevalence of neurofibromatosis 1 in German children at elementary school enrollment. *Arch Dermatol*. 2005;141(1):71–74.
6. Evans DG, Howard E, Giblin C, et al. Birth incidence and prevalence of tumor-prone syndromes: estimates from a UK family genetic register service. *Am J Med Genet A*. 2010;152A(2):327–332.
7. Ferner RE, Huson SM, Thomas N, et al. Guidelines for the diagnosis and management of individuals with neurofibromatosis 1. *J Med Genet*. 2007;44(2):81–88.
8. Huson SM, Harper PS, Compston DA. Von Recklinghausen neurofibromatosis. A clinical and population study in south-east Wales. *Brain*. 1988;111 (Pt 6):1355–1381.
9. Stewart DR, Korf BR, Nathanson KL, Stevenson DA, Yohay K. Care of adults with neurofibromatosis type 1: a clinical practice resource of the American College of Medical Genetics and Genomics (ACMG). *Genet Med*. 2018;20(7):671–682.
10. Ehara Y, Yamamoto O, Kosaki K, Yoshida Y. Natural course and characteristics of cutaneous neurofibromas in neurofibromatosis 1. *J Dermatol*. 2018;45(1):53–57.
11. Page PZ, Page GP, Ecosse E, Korf BR, Lepège A, Wolkenstein P. Impact of neurofibromatosis 1 on quality of life: a cross-sectional study of 176 American cases. *Am J Med Genet A*. 2006;140(18):1893–1898.
12. Draucker CB, Nutakki K, Varni JW, et al. The health-related quality of life of children, adolescents, and young adults with neurofibromatosis type 1 and their families: analysis of narratives. *J Spec Pediatr Nurs*. 2017;22:e12174.
13. Taylor LA, Lewis VL Jr. Neurofibromatosis type 1: review of cutaneous and subcutaneous tumor treatment on quality of life. *Plast Reconstr Surg Glob Open*. 2019;7(1):e1982.
14. Riccardi VM. A controlled multiphase trial of ketotifen to minimize neurofibroma-associated pain and itching. *Arch Dermatol*. 1993;129(5):577–581.
15. Ortonne N, Wolkenstein P, Blakeley JO, et al. Cutaneous neurofibromas: current clinical and pathologic issues. *Neurology*. 2018;91 (2 Suppl 1):S5–S13.
16. Louis DN, Perry A, Reifenger G, et al. The 2016 World Health Organization classification of tumors of the central nervous system: a summary. *Acta Neuropathol*. 2016;131(6):803–820.
17. Louis DN, Ohgaki H, Wiestler OD, et al. The 2007 WHO classification of tumours of the central nervous system. *Acta Neuropathol*. 2007;114(2):97–109.
18. Vidal CI, Armbrect EA, Andea AA, et al. Appropriate use criteria in dermatopathology: initial recommendations from the American Society of Dermatopathology. *J Am Acad Dermatol*. 2019 80(1):189–207.
19. Vidal CI, Armbrect EA, Andea AA, et al. Appropriate use criteria in dermatopathology: initial recommendations from the American Society of Dermatopathology. *J Cutan Pathol*. 2018;45(8):563–580.
20. Rosenberg AE. WHO classification of soft tissue and bone, fourth edition: summary and commentary. *Curr Opin Oncol*. 2013;25(5):571–573.
21. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009;42(2):377–381.