2017

Updated nomenclature for human and mouse neurofibromatosis type 1 genes

Corina Anastasaki
Washington University School of Medicine in St. Louis

Lu Q. Le
University of Texas Southwestern Medical Center

Robert A. Kesterson
University of Alabama at Birmingham

David H. Gutmann
Washington University School of Medicine in St. Louis

Corina Anastasaki, PhD
Lu Q. Le, MD, PhD
Robert A. Kesterson, PhD
David H. Gutmann, MD, PhD

Neurol Genet
2017;3:e169; doi: 10.1212/NXG.0000000000000169

UPDATED NOMENCLATURE FOR HUMAN AND MOUSE NEUROFIBROMATOSIS TYPE 1 GENES

OPEN

Neurofibromatosis type 1 (NF1; OMIM 162200) is one of the most common neurogenetic conditions, affecting 1 in 3,000 people worldwide. Characterized by a propensity to develop nervous system tumors, learning and behavioral deficits, and pigmentary abnormalities, NF1 is caused by a germline sequence alteration in the *NF1* gene (OMIM: 613113; chromosome 17q11.2). In addition, somatic *NF1* sequence changes have been reported in numerous other cancers,[1] extending their importance to malignancies occurring in the general population.

The *NF1* gene, comprising 57 exons, was initially identified, and its complete coding sequence was assembled over 25 years ago.[2] In the ensuing decades, 4 alternatively spliced exons (9a, 10a-2, 23a, and 48a) were discovered and added to a large number of exons with temporary numerical assignments (e.g., 10a2 or 23.2). This has led to considerable confusion and inconsistencies in clinical and scientific publications, which also has critical implications for *NF1* DNA sequence alteration reporting as part of patient management, genotype-phenotype correlations, and the interpretation of small animal models engineered with patient-specific sequence changes. To provide a unified annotation system for the *NF1* gene, the mouse and human genes were aligned and assembled using published sequences for sequential numbering and comparison.

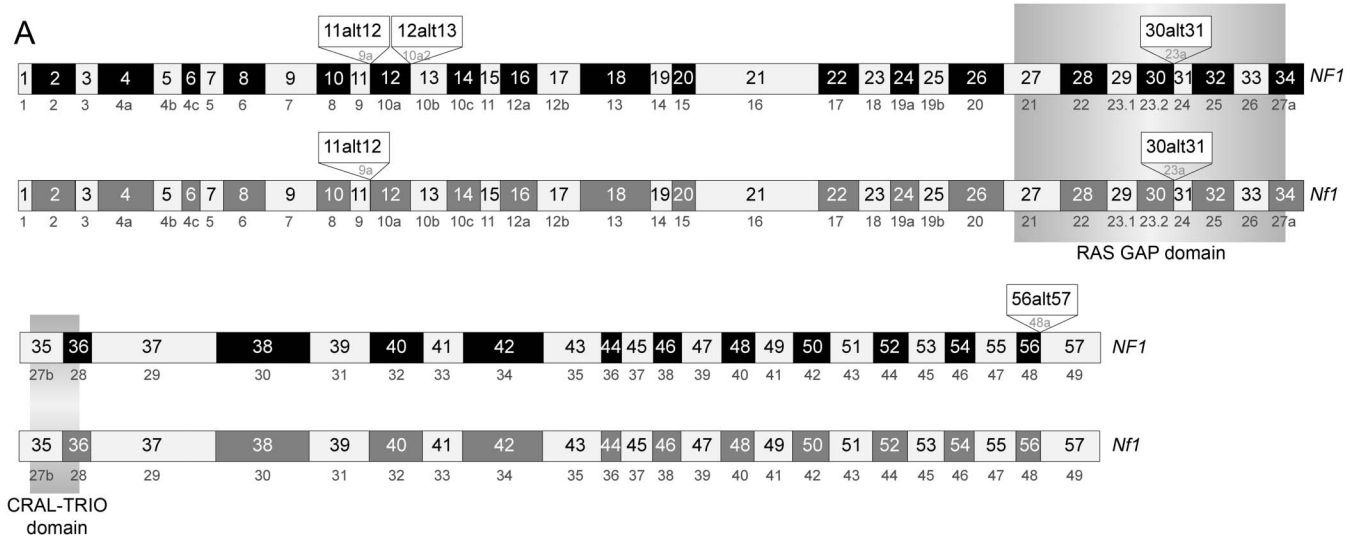
The full-length cDNA sequences of *NF1* (8457 nucleotides; figure e-1 at Neurology.org/ng) and *Nf1* (8463 nucleotides; figure e-2) genes were initially aligned using Nucleotide Basic Local Alignment Search Tool (BLAST) (NIH) optimized for megablast, revealing 92% sequence identity with 0.08% mismatches (figure e-2). Excluding the alternatively spliced exons, the ubiquitously expressed 57 exons were consecutively numbered (figure 1 and tables e1–e3). As such, the full-length human *NF1* transcript, NF1-002 (ENST00000356175.7), contains 57 exons and encodes 2818 amino acids, while the corresponding full-length mouse *Nf1* transcript, Nf1-003 (ENSMUST00000108251.8), comprises 57 exons and encodes 2,820 amino acids, with 2 additional amino acids encoded by exon 17. Alignment of the predicted amino acid sequences of the human (RefSeq NM_000267; UniProt P21359) and mouse (UniProt Q04690) full-length transcripts using protein BLAST revealed 98% amino acid sequence identity (figure e-3).

Next, we examined the previously published alternatively spliced exons: exon 9a contains a 30-nucleotide sequence encoding 10 amino acids and is predominantly expressed in postmitotic brain neurons,[3] whereas the muscle-specific exon 48a contains 54 nucleotides that encode 18 amino acids at the carboxyl terminus of the protein.[4] The 63-nucleotide exon 23a encodes 21 amino acids that are inserted within the RAS-GTPase-activating protein (GAP)-related domain of the *NF1* protein (neurofibromin), resulting in diminished RAS-GAP activity.[4] The most recently described alternatively spliced exon, exon 10a-2, contains 45 nucleotides and encodes 15 amino acids, is located at the amino terminus of neurofibromin, and this sequence has been hypothesized to direct intracellular membrane targeting by virtue of its predicted transmembrane domain.[5] Using published alternatively spliced exon sequences (figure e-1), BLAST searches were performed, and these alternatively spliced exons were renumbered and named according to their location (figure 1)—between exons 11–12 (11alt12; formerly 9a), exons 12–13 (12alt13; formerly 10a-2), exons 30–31 (30alt31; formerly 23a), and exons 56–57 (56alt57; formerly 48a). In this regard, we propose that the nucleotides (and the encoded amino acids) of these alternatively spliced exons be numbered 11alt12_1–30 (1–10), 12alt13_1–45 (1–15), 30alt31_1–63 (1–21), and 56alt57_1–54 (1–18). Possible new alternatively spliced exons should be similarly named based on their location.

11alt12 is 100% identical to both the cDNA and amino acid sequences of the predicted *Mus musculus* transcript variant X7 (mRNA XM_006532442.3), while 30alt31 is 98.4% identical to the cDNA and 95.2% identical to the amino acid sequence (containing a single conservative lysine-to-arginine change) of *Mus musculus Nf1* (ENSMUST00000071325.8) (figures e-4 and e-5). Notably, the 12alt13 and 56alt57 human sequences were not found in either mouse transcriptome or genome databases following

Supplemental data at Neurology.org/ng

Figure 1 **Proposed nomenclature for human *NF1* and mouse *Nf1* genes**



B

Exon #	Human Nucleotides	Human Amino acids	Mouse Nucleotides	Mouse Amino acids
1	1-60	1-20	1-60	1-20
2	61-204	21-68	61-204	21-68
3	205-288	69-97	205-288	69-96
4	289-479	98-160	289-479	97-160
5	480-586	160-195	480-586	160-196
6	587-654	196-218	587-654	196-218
7	655-730	219-244	655-730	219-244
8	731-888	244-296	731-888	244-296
9	889-1062	297-354	889-1062	297-353
10	1063-1185	355-395	1063-1185	354-395
11	1186-1260	396-420	1186-1260	396-420
12	1261-1392	421-464	1261-1392	421-464
13	1393-1527	465-508	1393-1527	465-509
14	1528-1641	509-546	1528-1641	510-547
15	1642-1721	547-574	1642-1721	548-574
16	1722-1845	574-615	1722-1845	574-615
17	1846-2001	616-667	1846-2007	616-669
18	2002-2251	668-751	2008-2257	670-753
19	2252-2325	751-775	2258-2331	753-777
20	2326-2409	776-803	2332-2415	778-805
21	2410-2850	804-950	2416-2856	806-952
22	2851-2990	951-997	2857-2996	953-999
23	2991-3113	997-1038	2997-3119	999-1040
24	3114-3197	1038-1066	3120-3203	1040-1068
25	3198-3314	1066-1105	3204-3320	1068-1107
26	3316-3496	1105-1166	3321-3502	1107-1168
27	3497-3708	1166-1236	3503-3714	1168-1238
28	3709-3870	1237-1290	3715-3876	1239-1292
29	3871-3974	1291-1325	3877-3980	1293-1327
30	3975-4110	1326-1670	3981-4116	1327-1372
31	4111-4269	1671-1423	4117-4275	1373-1425
32	4270-4367	1424-1456	4276-4373	1426-1458
33	4368-4514	1456-1505	4374-4520	1458-1507
34	4515-4661	1505-1554	4521-4667	1507-1556
35	4662-4772	1555-1591	4668-4778	1556-1593
36	4773-5205	1591-1735	4779-5211	1593-1737
37	5206-5546	1736-1849	5212-5552	1738-1851
38	5547-5749	1849-1917	5553-5755	1851-1919
39	5750-5943	1917-1981	5756-5949	1919-1893
40	5944-6084	1982-2028	5950-6090	1894-2030
41	6085-6364	2029-2122	6091-6370	2031-2124
42	6365-6579	2122-2193	6371-6585	2124-2195
43	6580-6641	2194-2214	6586-6647	2196-2216
44	6642-6756	2214-2252	6648-6762	2216-2254
45	6757-6858	2253-2286	6763-6864	2255-2288
46	6859-6999	2287-2333	6865-7005	2289-2335
47	7000-7126	2334-2376	7006-7132	2336-2378
48	7127-7258	2376-2420	7133-7264	2378-2422
49	7259-7394	2420-2465	7265-7400	2422-2467
50	7395-7552	2465-2518	7401-7558	2467-2520
51	7553-7675	2518-2559	7559-7681	2520-2561
52	7676-7806	2559-2602	7682-7812	2561-2604
53	7807-7907	2603-2636	7813-7913	2605-2638
54	7908-8050	2636-2684	7914-8056	2638-2686
55	8051-8097	2684-2699	8057-8103	2686-2701
56	8098-8314	2700-2772	8104-8320	2702-2774
57	8315-8457	2772-2818	8321-8463	2774-2820

(A) Schematic diagram illustrating the 57 exons of the human (top panel) and mouse (bottom panel) full-length transcripts, as well as the positions of the alternatively spliced exons (11alt12, 12alt13, 30alt31, and 56alt57). The proposed numbering system is denoted in each exon, while the prior numbering system is indicated below each exon. The RAS-GTPase-activating protein and the CRAL-TRIO domains are highlighted in gray. (B) Table detailing the nucleotides and amino acids spanning each of the 57 exons in human *NF1* and mouse *Nf1* genes. CRAI-TRIO domain = cellular retinaldehyde-binding protein and TRIO guanine exchange factor structural domain.

megablast, discontinuous megablast, or BLASTN searches (figures e-4 and e-5). Based on the absence of 12alt13 and 56alt57 sequences in the mouse, further studies will be required to define their roles in neurofibromin function. It should be noted that neither 12alt13 nor 56alt57 has been reported in other reference model organisms (rat or zebrafish; data not shown).

Taken together, we assembled and updated the human *NF1* and mouse *Nf1* sequences to facilitate greater consistency in both clinical reporting and basic research studies on NF1. The clarification of the nomenclature for the *NF1* gene is a necessary step to enable standardization of exon, nucleotide, and amino acid numbering, which is particularly important for reporting patient-specific DNA sequence variants. In addition, implementation of this proposed numbering system will provide better calibration of human genetic testing results with *Nf1* preclinical precision medicine models developed using genetically engineered mice[6] and induced pluripotent stem cells.[7]

From the Department of Neurology (C.A., D.H.G.), Washington University in St. Louis, MO; Department of Dermatology (L.Q. L.), University of Texas Southwestern Medical Center, Dallas; and Department of Genetics (R.A.K.), The University of Alabama at Birmingham.

Author contributions: C.A. performed the analyses, assembled the figures and tables, and helped write the manuscript. L.Q.L. and R.A.K. were involved in the design of the study and edited the manuscript. D.H.G. wrote the manuscript with C.A. and oversaw the study at all stages.

Study funding: D.H.G. is supported by a Research Program Award from the National Institute of Neurological Disorders and Stroke (1-R35-NS097211-01). The authors also acknowledge support from The Giorgio Foundation.

Disclosure: C. Anastasaki and L.Q. Le report no disclosures. R.A. Kesterson has served on the editorial board of Frontiers in Genetics *and has received research support from UAB Hepato/Renal Fibrocystic Diseases Core Center, UAB Comprehensive Cancer Center, UAB Center for Nutrition and Obesity Research, UAB Rheumatic Disease Core Center, UAB Diabetes Research Center, and the Giorgio Foundation. D.H. Gutmann has served on scientific advisory boards for the American Association for Cancer Research and the National Institute for Neurological Disorders and Stroke Advisory Council; has served on the editorial boards of* Glia and Familial Cancer*; holds patents for Identification of the NF1 gene and protein and Identification of NF1 protein as mTOR regulator; has received research support from the US Army Department of Defense, NCI Tumor Microenvironment Network, NIH, the Giorgio Foundation, and Children's Tumor Foundation; and receives royalty payments from the University of Michigan. Go to Neurology.org/ng for full disclosure forms. The Article Processing Charge was funded by the authors.*

Received February 10, 2017. Accepted in final form April 20, 2017.

Correspondence to Dr. Gutmann: gutmannd@wustl.edu

1. Patil S, Chamberlain RS. Neoplasms associated with germline and somatic NF1 gene mutations. Oncologist 2012;17:101–116.
2. Marchuk DA, Saulino AM, Tavakkol R, et al. cDNA cloning of the type 1 neurofibromatosis gene: complete sequence of the NF1 gene product. Genomics 1991;1:931–940.
3. Geist RT, Gutmann DH. Expression of a developmentally-regulated neuron-specific isoform of the neurofibromatosis 1 (NF1) gene. Neurosci Lett 1996;211:85–88.
4. Gutmann DH, Andersen LB, Cole JL, Swaroop M, Collins FS. An alternatively-spliced mRNA in the carboxy terminus of the neurofibromatosis type 1 (NF1) gene is expressed in muscle. Hum Mol Genet 1993;2:989–992.
5. Kaufmann D, Müller R, Kenner O, et al. The N-terminal splice product NF1-10a-2 of the NF1 gene codes for a transmembrane segment. Biochem Biophys Res Commun 2002;294:496–503.
6. Li K, Turner AN, Chen M, et al. Mice with missense and nonsense NF1 mutations display divergent phenotypes compared with human neurofibromatosis type I. Dis Model Mech 2016;9:759–767.
7. Anastasaki C, Woo AS, Messiaen LM, Gutmann DH. Elucidating the impact of neurofibromatosis-1 germline mutations on neurofibromin function and dopamine-based learning. Hum Mol Genet 2015;24:3518–3528.

Updated nomenclature for human and mouse neurofibromatosis type 1 genes

Corina Anastasaki, Lu Q. Le, Robert A. Kesterson, et al.

Neurol Genet 2017;3;
DOI 10.1212/NXG.0000000000000169