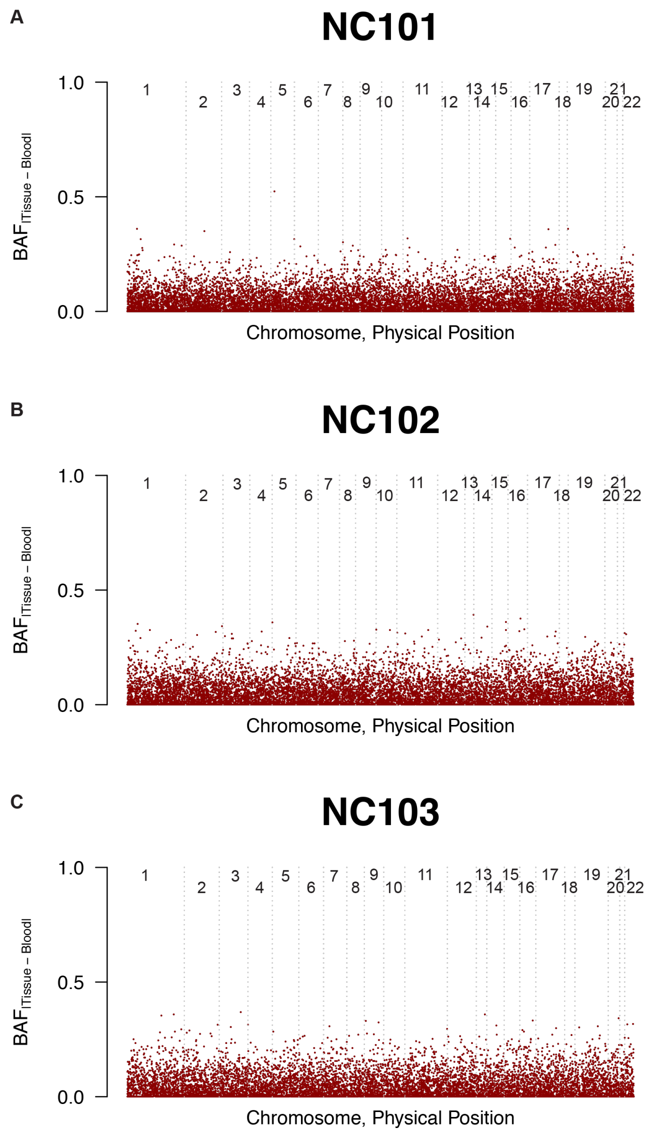
The American Journal of Human Genetics, Volume *98*

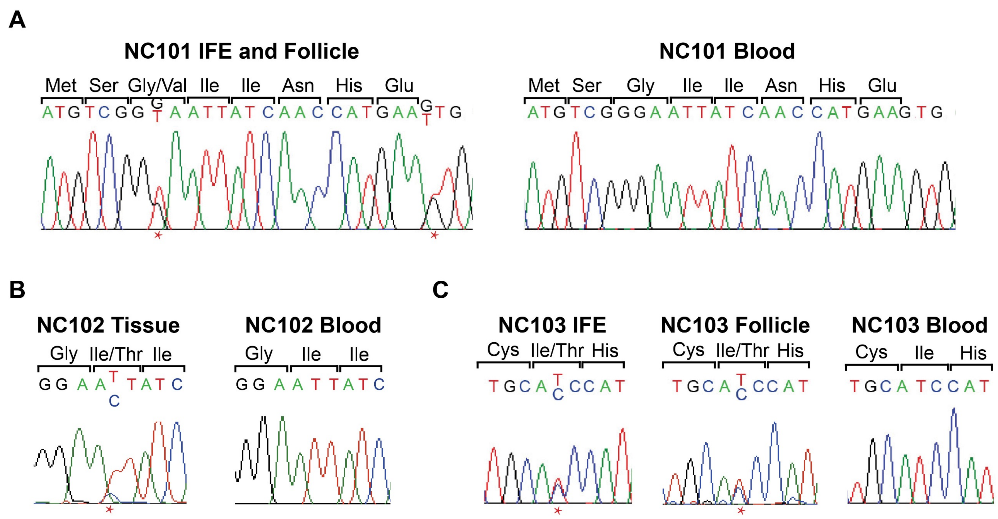
Supplemental Data

Somatic Mutations in *NEK9* Cause Nevus Comedonicus

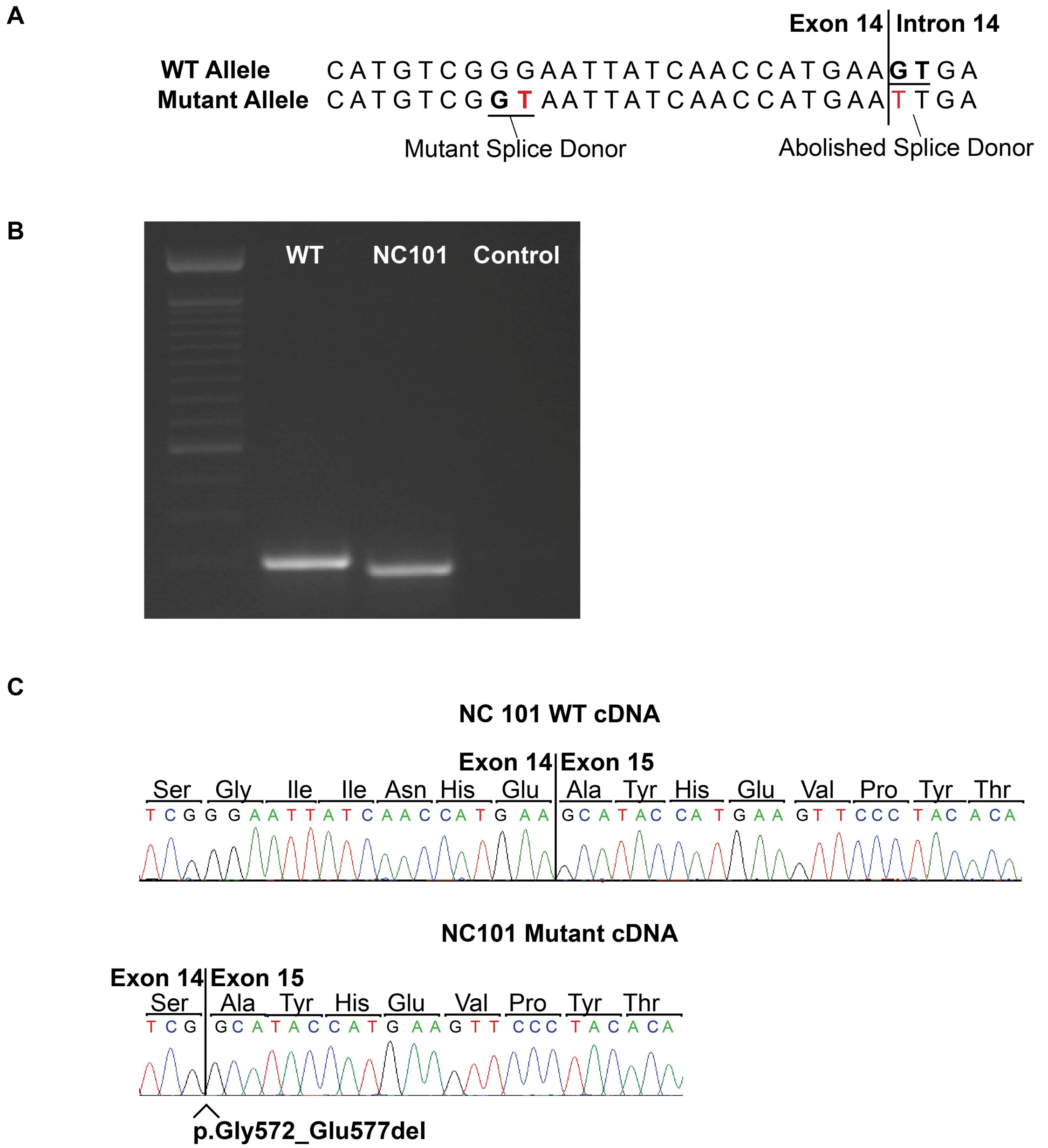
Jonathan L. Levinsohn, Jeffrey L. Sugarman, Yale Center for Mendelian Genomics, Jennifer M. McNiff, Richard J. Antaya, and Keith A. Choate



Supplemental Figure 1. NC lesions demonstrate no regions of loss of heterozygosity. Difference of B-allele frequency between tissue and blood is plotted against genomic position, demonstrating no regions with altered B-allele frequency. There is no evidence of large segments of loss of heterozygosity in any of the samples.



Supplemental Figure 2. Sanger sequencing confirms somatic mutations identified via whole exome sequencing. Sanger sequencing of PCR amplified gDNA isolated from laser-captured IFE and follicle tissue for NC101 (A) confirms two mutations. One mutation is at p.Gly572Val, creating a cryptic splice donor site, and the other mutation destroys the normal splice donor site, leading to deletion of Gly572 through Glu577 (Figure S3). These mutations are absent in blood. Sanger sequencing of NC102 gDNA also confirms the p.Ile573Thr mutation is present in DNA isolated from whole affected tissue (B), but is absent in blood. Sanger sequence of laser-captured IFE and follicle tissue NC103 (C) confirms a somatic p.Ile167Thr mutation in IFE and in follicle.



Supplemental Figure 3. Analysis of *NEK9* splicing mutation in NC101 tissue reveals a 6 amino acid deletion. Two mutations are present in NC101 tissue in *cis*. The first mutation appears to abolish a splice donor, and the second mutation creates a potential novel upstream splice donor (A). To experimentally assess the consequence of this mutant allele upon encoded RNA, high-fidelity Advantage II long range PCR (Clontech) of genomic DNA from NC101 was used to create amplicons consisting of exons 13-15 including flanking intronic regions. This amplicon was TA-cloned (Invitrogen). Clones were sequenced to confirm that they contained the mutant allele without additional mutations introduced by PCR, and were then cloned into pCDNA 3.1 using Kpn1 and Not1. WT and mutant plasmids were transfected into PLC cells. 48 hours after transfection, RNA was isolated using the RNEasy kit (QIAGEN). cDNA was generated using the iScript kit (BD Biosciences). PCR was then performed and products were run on a 1% agarose gel (B) demonstrating that NC101 mutant allele cDNA was shorter than the WT allele. Sanger sequencing of these products revealed that the NC101 mutant lacked 18 nucleotides encoding amino acids ‘Gly,Ile,Ile,Asn,His,Glu’ which fall within the RCC1 domain (C).

```
HUMAN_NEK9     MSVLGEYERHCDSINSDFGSESGGCGDSSPGPSASQGPRAGGGAAEQEELHYIPIRVLGR  60
RHESUS_NEK9    MSVLGEYERHCDSINSDFGSESGGCGDSSPGPSASQGPRAGGGAAEQEELHYIPIRVLGR  60
MOUSE_NEK9     MSVLGEYERHCDSINSDFGSESGGGGDSGPGPSAVPGPRAGGGAAEQEELHYIPIRVLGR  60
RAT_NEK9       MSVLGEYERHCDSINSDFGSESGGGGDSGSGPSASPGPRAGG-AAEQEELHYIPIRVLGR  59
COW_NEK9       MSVLGEYERHCDSLNSDFGSESGGGGDSGPGPSAGPVPRASGGAAEQEELHYIPIRVLGR  60
XENOPUS_NEK9   MSALGRYDRHCDSINSDFGDSVRSCG------------------PEQEELHYIPIRVLGH  42

HUMAN_NEK9     GAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNIIAYYNHFMD  120
RHESUS_NEK9    GAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNIIAYYNHFMD  120
MOUSE_NEK9     GAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNIIAYYNHFMD  120
RAT_NEK9       GAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNIIAYYNHFMD  119
COW_NEK9       GAFGEATLYRRTEDDSLVVWKEVDLTRLSEKERRDALNEIVILALLQHDNIIAYYNHFMD  120
XENOPUS_NEK9   GAYGEATLYRRTEDDSLVVWKEVGLARLSEKERRDALNEIVILSLLQHDNIIAYYNHFLD  102

                                                      * NC103
HUMAN_NEK9     NTTLLIELEYCNGGNLYDKILRQKDKLFEEEMVVWYLFQIVSAVSCIHKAGILHRDIKTL  180
RHESUS_NEK9    NTTLLIELEYCNGGNLYDKILRQKDKLFEEEMVVWYLFQIVSAVSCIHKAGILHRDIKTL  180
MOUSE_NEK9     NTTLLIELEYCNGGNLYDKILRQKDKLFEEEMVVWYLFQIVSAVSCIHKAGILHRDIKTL  180
RAT_NEK9       NTTLLIELEYCNGGNLYDKILRQKDKLFEEEMVVWYLFQIVSAVSCIHKAGILHRDIKTL  179
COW_NEK9       NTTLLIELEYCNGGNLYDKILRQKDKLFEEEMVVWYLFQIVSAVSCIHKAGILHRDIKTL  180
XENOPUS_NEK9   SNTLLIELEYCNGGNLFDKIVRQKAQLFQEEMVLWYLFQIVSAVSCIHRAGILHRDIKTL  162

HUMAN_NEK9     NIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNFKSDIWAVGCV  240
RHESUS_NEK9    NIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNFKSDIWAVGCV  240
MOUSE_NEK9     NIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNFKSDIWAVGCV  240
RAT_NEK9       NIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNFKSDIWAVGCV  239
COW_NEK9       NIFLTKANLIKLGDYGLAKKLNSEYSMAETLVGTPYYMSPELCQGVKYNFKSDIWAVGCV  240
XENOPUS_NEK9   NIFLTKANLIKLGDYGLAKQLSSEYSMAETCVGTLYYMSPELCQGVKYSFKSDIWAVGCV  222

HUMAN_NEK9     IFELLTLKRTFDATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLDQDPEQRPTAD  300
RHESUS_NEK9    IFELLTLKRTFDATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHSCLDQDPEQRPTAD  300
MOUSE_NEK9     IFELLTLKRTFDATNPLNLCVKIVQGIRAMEVDSSQYSLELIQLVHACLDQDPEQRPAAD  300
RAT_NEK9       IFELLTLKRTFDATNPLNLCVKIVQGIRAMEVDSSQYSLGLIQLVHACLDQDPERRPTAD  299
COW_NEK9       IFELLTLKRTFDATNPLNLCVKIVQGIRAMEVDSSQYSLELIQMVHACLDQDPEQRPTAD  300
XENOPUS_NEK9   LYELLTLTRTFDATNPLNLCVKIVQGNWAMGLDNSVYSQDLIEVVHACLEQDPEKRPTAD  282

HUMAN_NEK9     ELLDRPLLRKRRREMEEKVTLLNAPTKRPRSSTVTEAPIAVVTSRTSEVYVWGGGKSTPQ  360
RHESUS_NEK9    ELLDRPLLRKRRREMEEKVTLLNAPTKRPRSSTVTEAPIAVVTSRTSEVYVWGGGKSTPQ  360
MOUSE_NEK9     ALLDLPLLRTRRREMEEKVTLLNAPTKRPRSSTVTEAPIAVVTSRTSEVYVWGGGKSTPQ  360
RAT_NEK9       ALLDLPLLRKRRREMEEKVTLLNAPTKRPRSSTVTEAPIAVVTSRTSEVYVWGGGKSTPQ  359
COW_NEK9       ELLDRPLLRKRRREMEEKVTLLNAPTKRPRSSTVTEAPIAVVTSRTSEVYVWGGGKSTPQ  360
XENOPUS_NEK9   EILKMPILSWRRRDMEEKVSMLNRSNKKPRTGTVTEAPIAVVTSRSSEVYVWGGGKTTPQ  342
```

```
HUMAN_NEK9     KLDVIKSGCSARQVCAGNTHFAVVTVEKELYTWVNMQGGTKLHGQLGHGDKASYRQPKHV 420
RHESUS_NEK9    KLDVIKSGCSARQVCAGNTHFAVVTVEKELYTWVNMQGGTKLHGQLGHGDKASYRQPKHV 420
MOUSE_NEK9     KLDVIKSGCSARQVCAGNTHFAVVTVEKELYTWVNMQGGTKLHGQLGHGDKASYRQPKHV 420
RAT_NEK9       KLDVIKSGCSARQVCAGNTHFAVVTVEKELYTWVNMQGGTKLHGQLGHGDKASYRQPKHV 419
COW_NEK9       KLDVIKSGCSARQVCAGNTHFAVVTVEKELYTWVNMQGGTKLHGQLGHGDKASYRQPKHV 420
XENOPUS_NEK9   KLDVFKGGCRARQVCAGDAHFAVVTVEKELYTWVNMQGGSKLHGQLGHGDRASYRQPKHV 402
```

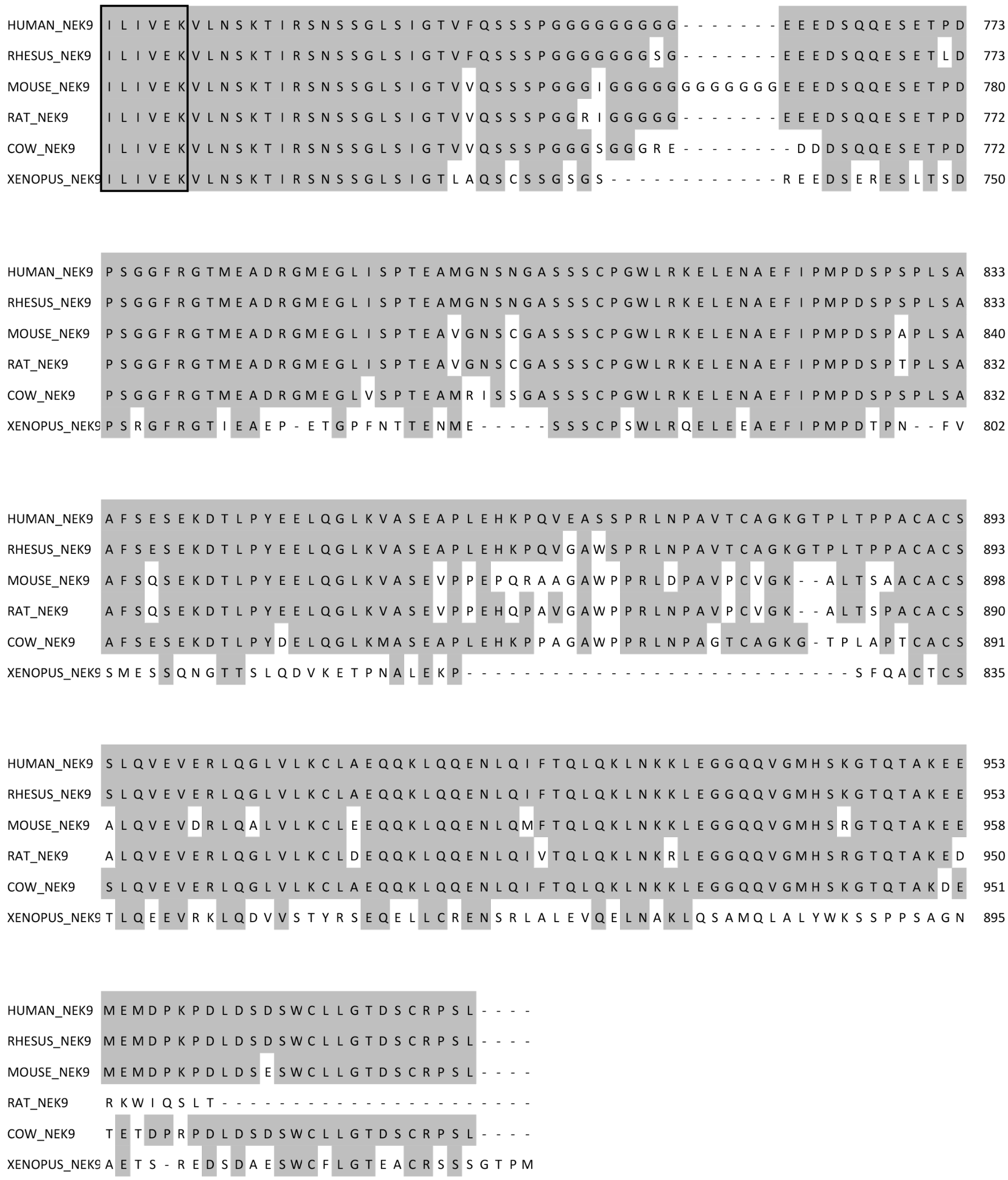
```
HUMAN_NEK9     EKLQGKAIRQVSCGDDFTVCVTDEGQLYAFGSDYYGCMGVDKVAGPEVLEPMQLNFFLSN 480
RHESUS_NEK9    EKLQGKAIHQVSCGDDFTVCVTDEGQLYAFGSDYYGCMGVDKVAGPEVLEPMQLNFFLSN 480
MOUSE_NEK9     EKLQGKAIHQVSCGDDFTVCVTDEGQLYAFGSDYYGCMGVDKVSGPEVLEPMQLNFFLSN 480
RAT_NEK9       EKLQGKAIHQVSCGDDFTVCVTDEGQLYAFGSDYYGCMGVDKVSGPEVLEPMQLNFFLSN 479
COW_NEK9       EKLQGKAIRQVSCGDDFTVCVTDEGQLYAFGSDYYGCMGVDKVAGAEVLEPMQLDFFLSN 480
XENOPUS_NEK9   EKLQGKSVQQVSCGSDFTVCISDEGQLYSFGSDYYGCLGVNQSAGAEVLEPLLVDFFLNE 462
```

```
HUMAN_NEK9     PVEQVSCGDNHVVVLTRNKEVYSWGCGEYGRLGLDSEEDYYTPQKVDVPKALIIVAVQCG 540
RHESUS_NEK9    PVEQVSCGDNHVVVLTRNKEVYSWGCGEYGRLGLDSEEDYYTPQKVDVPKALIIVAVQCG 540
MOUSE_NEK9     PVEQVSCGDNHVVVLTRNKEVYSWGCGEYGRLGLDSEEDYYTPQRVDVPKALIIVAVQCG 540
RAT_NEK9       PVEQVSCGDNHVVVLTRNKEVYSWGCGEYGRLGLDSEEDYYTPQRVDVPKALIIVAVQCG 539
COW_NEK9       PVEQVSCGDNHVVVLTRNKEVYSWGCGEYGRLGLDSEEDYYTPQKVDVPKALIIVAVQCG 540
XENOPUS_NEK9   PVEQVSCGDSHIIALTRSKCVYSWGCGEYGRLGLDSEDDVYSPQKVEVQRDLCIVNVCCG 522
```

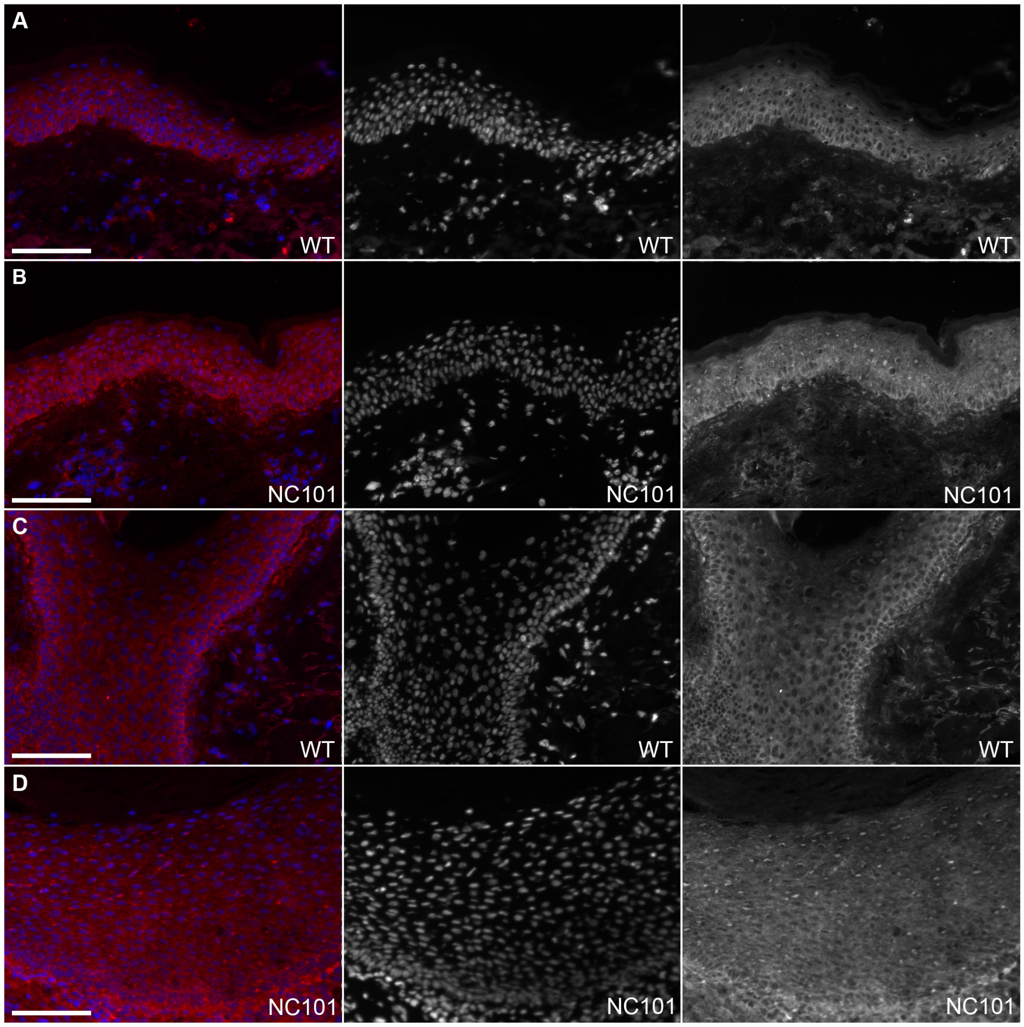
```
                                                   * NC102
                                                ****** NC101
HUMAN_NEK9     CDGTFLLTQSGKVLACGLNEFNKLGLNQCMSGIINHEAYHEVPYTTSFTLAKQLSFYKIR 600
RHESUS_NEK9    CDGTFLLTQSGKVLACGLNEFNKLGLNQCMSGIINHEAYHEVPYTTSFTLAKQLSFYKIR 600
MOUSE_NEK9     CDGTFLLTQSGKVLACGLNEFNKLGLNQCMSGIINHEAYHEVPYTTSFTLAKQLSFYKIR 600
RAT_NEK9       CDGTFLLTQSGKVLACGLNEFNKLGLNQCMSGIINHEAYHEVPYTTSFTLAKQLSFYKIR 599
COW_NEK9       CDGTFLLTQSGKVLACGLNEFNKLGLNQCMSGIINHEAYHEVPYTTSFTLAKQLSFYKIR 600
XENOPUS_NEK9   SDGSFLLTLTGKVLACGLNEHNKLGLNQYTAGIINHEAFQEVPYTTSLTLAKQLSFYKIR 582
```

```
HUMAN_NEK9     TIAPGKTHTAAIDERGRLLTFGCNKCGQLGVGNYKKRLGINLLGGPLGGKQVIRVSCGDE 660
RHESUS_NEK9    TIAPGKTHTAAIDERGRLLTFGCNKCGQLGVGNYKKRLGINLLGGPLGGKQVIRVSCGDE 660
MOUSE_NEK9     TIAPGKTHTAAIDERGRLLTFGCNKCGQLGVGNYKKRLGINLLGGPLGGKQVIRVSCGDE 660
RAT_NEK9       TIAPGKTHTAAIDERGRLLTFGCNKCGQLGVGNYKKRLGINLLGGPLGGKQVIRVSCGDE 659
COW_NEK9       TIAPGKTHTAALDERGRLLTFGCNKCGQLGVGNYKKRLGINLLGGPLGGKQVIRVSCGDE 660
XENOPUS_NEK9   SISPGRTHTAAIDERGRLLTFGSNKCGQLGVGDYRKHLGINLLGGPLGGKQVIRVSCGDE 642
```

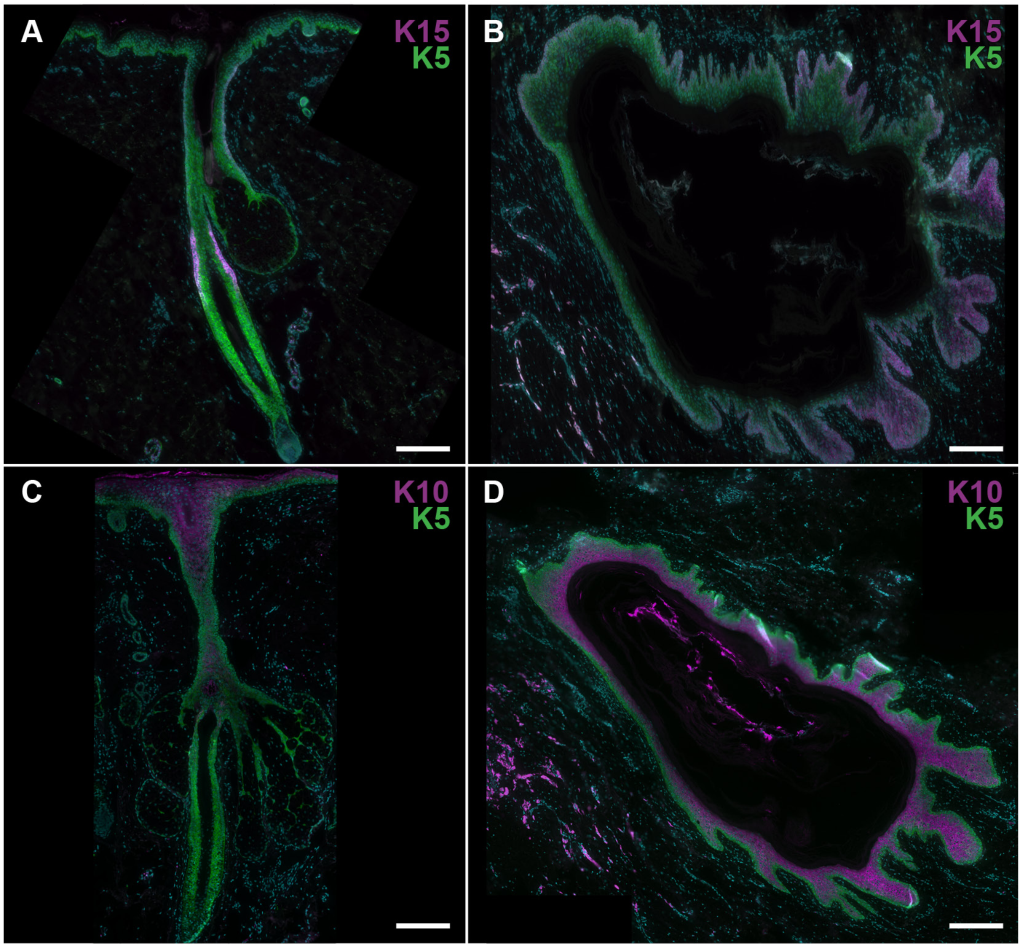
```
HUMAN_NEK9     FTIAATDDNHIFAWGNGGNGRLAMTPTERPHGSDICTSWPRPIFGSLHHVPDLSCRGWHT 720
RHESUS_NEK9    FTIAATDDNHIFAWGNGGNGRLAMTPTERPHGSDICTSWPRPIFGSLHHVPDLSCRGWHT 720
MOUSE_NEK9     FTIAATDDNHIFAWGNGGNGRLAMTPTERPHGSDICTSWPRPIFGSLHHVPDLSCRGWHT 720
RAT_NEK9       FTIAATDDNHIFAWGNGGNGRLAMTPTERPHGSDICTSWPRPIFGSLHHVPDLSCRGWHT 719
COW_NEK9       FTIAATDDNHIFAWGNGGNGRLAMTPTERPHGSDICTSWPRPIFGSLHHVPDLSCRGWNT 720
XENOPUS_NEK9   FTIAATADNHIFAWGNGGNGRLAMTPNERPQGSDICTSWPRPIFGSLHHVTDLSCRGWHT 702
```



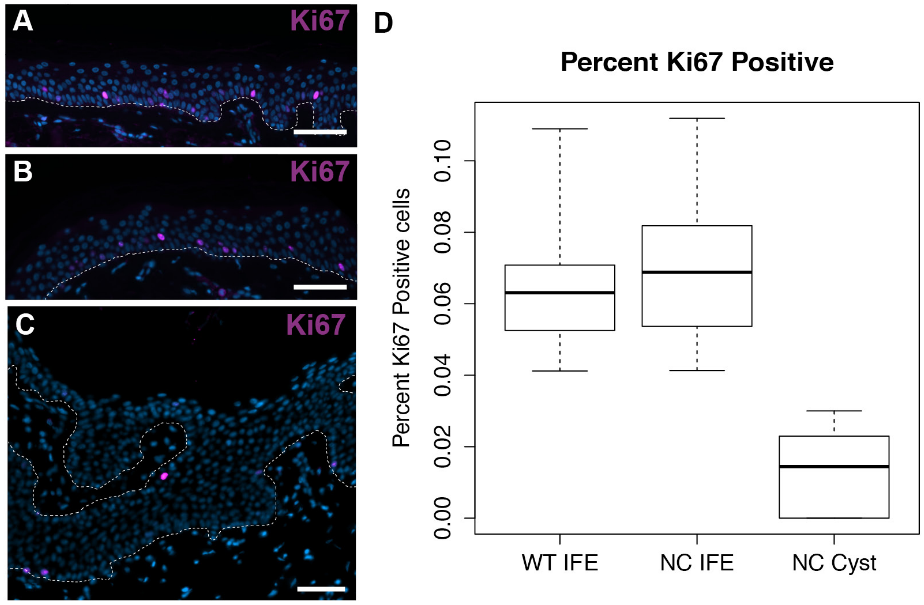
Supplemental Figure 4. NC101, NC102 and NC103 mutations are at conserved residues. Human, rhesus, mouse, rat, cow and *Xenopus NEK9* orthologs were aligned using CLUSTAL-omega. The amino acids affected by NC101, NC102 and NC103 mutations are indicated by a red asterisk above the amino acid(s) involved. All mutations affect amino acids that are highly conserved across vertebrate species. The kinase domain (residues 52-308) and RCC1 domain (residues 388-726) are outlined in black.



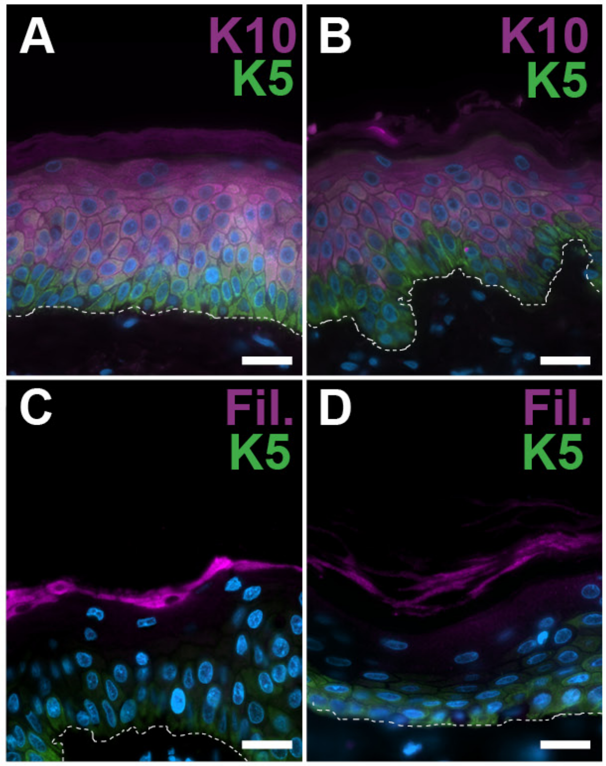
Supplemental Figure 5. NEK9 shows cytoplasmic localization in wild-type and NC epidermis. Frozen lesional NC tissue and wild-type (WT) tissue from margins of surgical excisions were fixed with 1:1 methanol:acetone at -20 C. Rabbit Anti-NEK9 antibody (Abcam; ab138488) was used at 1:250 dilution. Immunolocalization studies show that NEK9 localizes to the cytoplasm in IFE (A) and hair follicles (C) in wild-type tissue, and shows cytoplasmic localization in NC101 IFE (B) and cysts (D). Scale = 100 μm.



Supplemental Figure 6. Expansion of K15 localization and ectopic K10 immunolocalization in NC comedones and cysts. Stitched images of normal and NC tissue further illustrate the defect in differentiation found in NC. In normal hair follicles, K15 is found localized within the bulge region (A). However in NC, we see ectopic K15 staining of deep dermal cysts, (B). K10 is restricted to the IFE and superficial regions of the wild-type hair follicles (C), however we see suprabasal localization of K10 in NC follicles and deep dermal cysts (D). Scale = 200 µm.



Supplemental Figure 7. Ki67 immunolocalization reveals no evidence of hyperproliferation in NC tissue. To assess proliferation within NC lesions, Ki67 (Abcam; 15580, 1:300 dilution), positive cells were counted in twenty 20x fields of view and normalized to total number of basal keratinocytes. T-tests were performed to compare the NC IFE, WT IFE, and NC cyst tissue. T-test of Ki67 staining revealed similar number of positive cells in WT IFE (A) and NC IFE (B) tissue, p-value = 0.459 (D). Cystic regions of NC tissue demonstrated fewer Ki67 positive cells than either WT or NC IFE (C,D), p-values < 10^{-8}, T-test. Scale = 50 µm.



Supplemental Figure 8. NC IFE shows features of normal differentiation. Keratin 10 localizes to supra basal layers of wilt-type tissue (A) (Santa Cruz; sc-53252, 1:200 dilution) (A) and is also suprabasal in NC103 IFE (B). Similar localization of filaggrin (Abcam; ab17808, 1:400 dilution) is seen in wild-type (C) and NC103 IFE (D). Keratin 5 staining marks the basal layer of the IFE. Dermal-epidermal junction is marked by dashed white line. Scale bar = 20 µm.

Sample	Mean Coverage	Bases Covered >8x	Bases Covered >20x	Mean Read Length
NC101 Tissue	189.9x	98%	97%	74 bases
NC101 Blood	107.9x	97%	93%	74 bases
NC102 Tissue	93.4x	98%	94%	74 bases
NC102 Blood	86.9x	97%	88%	74 bases
NC103 Tissue	119.5x	97%	94%	74 bases
NC103 Blood	86.6x	97%	93%	74 bases

Supplemental Table 1. Whole exome sequencing coverage. Whole exome paired-end 74bp sequencing of tissue and blood was performed for three samples. DNA was sheared and barcoded, followed by capture with Roche EZ exome V3 capture probes. Sequencing was performed using the Illumina HiSeq 2500. Blood samples were run at 6 samples per lane, while tissue samples were run at 4 samples per lane. Sequence was aligned to hg19 using BWA-MEM. Reads were then trimmed. PCR duplicates removed using Picard and BAM files were calibrated with GATK. In all samples, >97% of coding region bases are covered >8x yielding sufficient coverage for analysis.

SSNVs Called with Filter	NC101	NC102	NC103
All Called	16	50	22
Within Exons/ Splice Sites	9	21	22
Nonsynonymous	7	13	15
> 0.1% Prevelance in ExAC	3	6	13
Fisher Test p-value > 0.1%	2*	1*	1*

Supplemental Table 2. SSNV filtering results. To identify tissue-specific mutations, a Perl script was used in tandem with MuTect. A Fisher Exact test was performed to determine non-reference read enrichment in tissue, and SSNVs with p-value greater than 1×10^{-3} were excluded; genome wide significance is 1.7×10^{-6} after Bonferroni correction. The smallest p-values of the Fisher test for damaging SSNVs not within *Nek9* were 0.25, 0.02 and 0.12 in NC101, NC102 and NC103, respectively. To remove non-damaging SSNVs, we filtered to exclude synonymous mutations and intronic variants. SSNVs reported at greater than 1% of the population in the ExAC control database were filtered. SSNVs were then inspected on Integrative Genome Viewer (IGV) to ensure correct mapping. The remaining 4 *NEK9* mutations were not found within ExAC, 1000 Genomes and NHLBI Exome Variant Server, and dbSNP control data sets. Mutations found via WES were validated with Sanger sequencing, and confirmed via laser capture microscopy in NC101 and NC103, for which tissue was available (Figure S2). Asterisk indicates that all of these mutations were within *NEK9.*