Table S1: Primers for real-time PCRs of RA signaling downstream genes

Gene name	Forward	Reverse
Rarb	acccacctcccagtggat	gctggtactctgtgtctcgatg
Crabp1	aaggtcggagagggcttc	tgtgcagtgaatcttgttctca
Wnt11	caggatcccaagccaataaa	tccagggaggcacgtaga
Cdh2	gccatcatcgctatccttct	ccgtttcatccataccacaaa
Foxa1	cgcaggtacgagtttcgtg	cgtcctctccccatttgtc
Hoxb1	aagagaaacccacctaagacagc	tgaagtttgtgcggagacc
Casp3	gaggctgacttcctgtatgctt	aaccacgacccgtccttt
Pou5f1	gttggagaaggtggaaccaa	ctccttctgcagggctttc
GFP	cttcagctaccgctacgagg	gctgccatccagatcgttat
Gapdh	aagagggatgctgcccttac	ccattttgtctacgggacga

Table S2: Summary for significant association of common variants with risk of NTDs

Gene symbol	SNP ID	Genomic position	Ref allele	Alt allele	HOMR/HET/HOMA in NTD	HOMR/HET/HOMA in controls	HWE in controls	OR(95% CI)	*P* value
CRABP2	rs12039622	chr1:156669917	T	C	284/67/4	186/28/1	1	1.60 (1.00-2.57)	0.049
CYP26B1	rs917896	chr2:72374435	C	T	0/45/269	0/39/137	0.14	1.70 (1.06-2.74)	0.029
ALDH1A2	rs4646579	chr15:58329528	G	C	194/119/16	105/71/20	0.16	0.45 (0.23-0.89)	0.022
ALDH1A2	rs4238328	chr15:58338651	G	A	228/109/18	115/83/13	0.86	0.67 (0.47-0.94)	0.022

SNP ID is referred as dbSNP in NCBI.

Genomic position is referenced from the human genome in UCSC (hg19/GRCh37).

HOMR: homozygous reference genotype; HET: heterozygous genotypes; HOMA: homozygous alterative genotype.

HWE: Hardy-Weinberg equilibrium.

OR: odds ratio

Table S3: Constraint metrics of targeted genes in the present cohort

Gene symbol	Constraint from ExAC	Expected no. variants	Observed no. variants	Constraint Metrics
ALDH1A2	**Synonymous**	78.6	90	z = -0.80
	Missense	176.4	141	z =1.30
	LoF	19.1	4	pLI = 0.52
	CNV	8.6	32	z = -1.62
CYP26A1	**Synonymous**	109	98	z = 0.65
	Missense	223.9	168	z =1.83
	LoF	15.5	9	pLI = 0.00
	CNV	4.0	2	z = 0.38
CYP26B1	**Synonymous**	117.1	116	z = 0.07
	Missense	239.3	174	z =2.06
	LoF	9.4	0	pLI = 0.95
	CNV	3.0	2	z = 0.20
CRABP1	**Synonymous**	34.4	28	z = 0.67
	Missense	61.5	35	z =1.65
	LoF	5.2	2	pLI = 0.13
	CNV	6.2	1	z = 0.94
CRABP2	**Synonymous**	25.5	27	z = -0.19
	Missense	47.6	48	z =-0.03
	LoF	6.7	4	pLI = 0.01
	CNV	3.6	0	z = 0.80
RARA	**Synonymous**	87.9	101	z = -0.87
	Missense	182.1	83	z =3.59
	LoF	12.8	1	pLI = 0.93
	CNV	4.6	1	z = 0.70

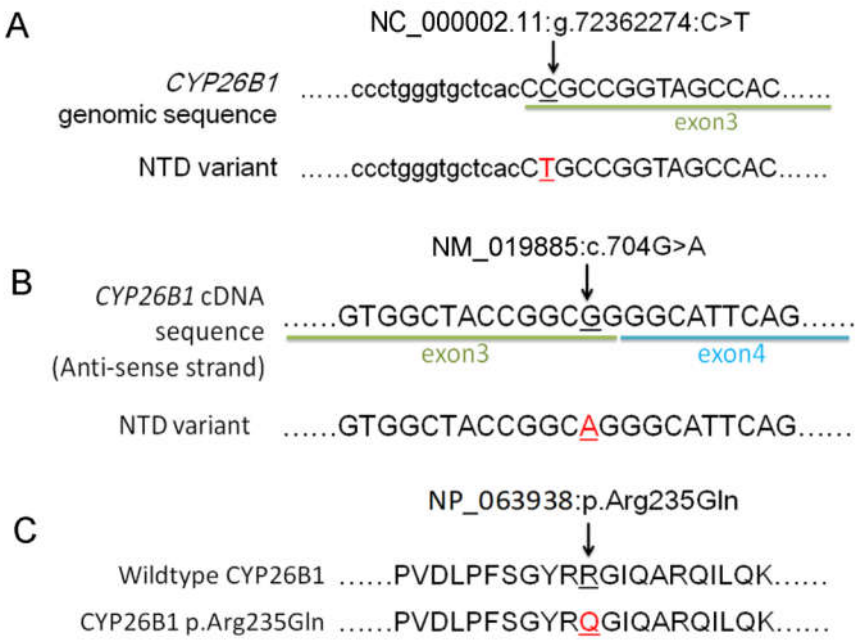

FIGURE S1: Schematic of *CYP26B1* variant NC_000002.11:g.72362274:C>T NP_063938:p.(Arg235Gln) splicing and missense mutation

A) Genomic DNA sequence of the *CYP26B1* exon 3/intron region showing the reference sequence and NTD-variant sequence at g.72362274:C>T p.(Arg235Gln) from GRCh37/hg19. The base change in the NTD variant from C to T is shown in red. Upper case letters indicate exon 3 sequence and lower case letters indicate intronic sequence. The cDNA (B; anti-sense strand) and protein (C) sequences of reference and NTD-specific g.72362274:C>T p.(Arg235Gln) variant are also shown. The red font "A" is the altered allele from the reference "G" in the cDNA sequence; the red "Q" indicates predicted amino acid change from arginine (R) to glutamine (Q).

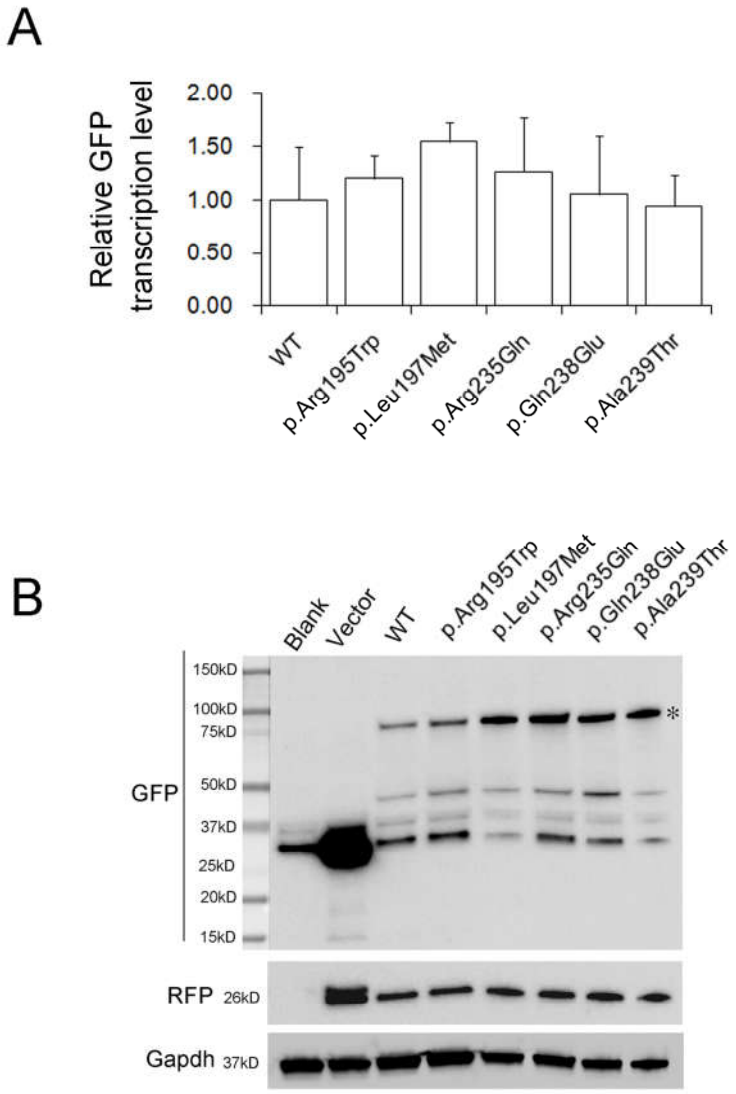

FIGURE S2: Transfection efficiency in NE-4C cells

A) Real-time PCR results for *GFP* expression from the *CYP26B1-GFP fusion* show that the transfection efficiencies between WT and mutants are not significantly different. All results in mutants were normalized by WT. B) Western blot image showing proteins expressed following transfection of NE-4C cells with WT or mutant CYP26B1-GFP fusion plasmids and RFP transfection control plasmid. Asterisk on right indicates the expected size of the CYP26B1-GFP fusion protein. Gapdh was used as a control for protein loading.

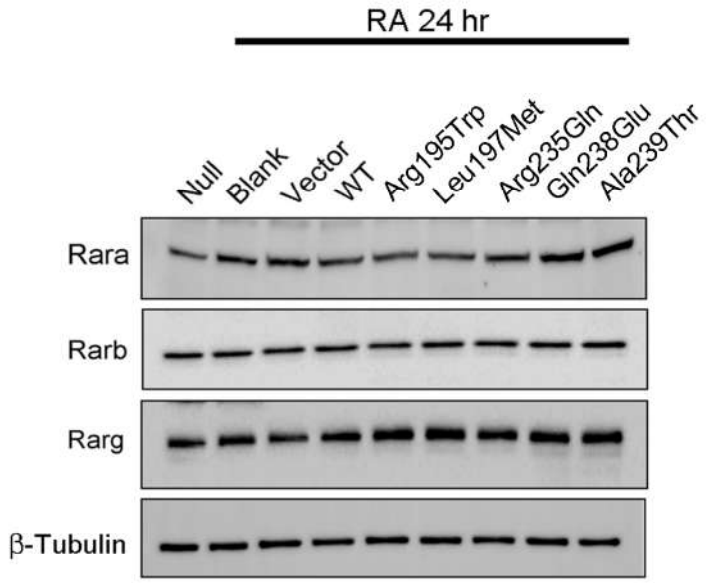

FIGURE S3: Western blots indicate the relative protein expression of RA receptors 24 hours after RA-treatment of NE-4C cells transfected with *CYP26B1* WT or mutant plasmids.