Supplementary Figures and Tables

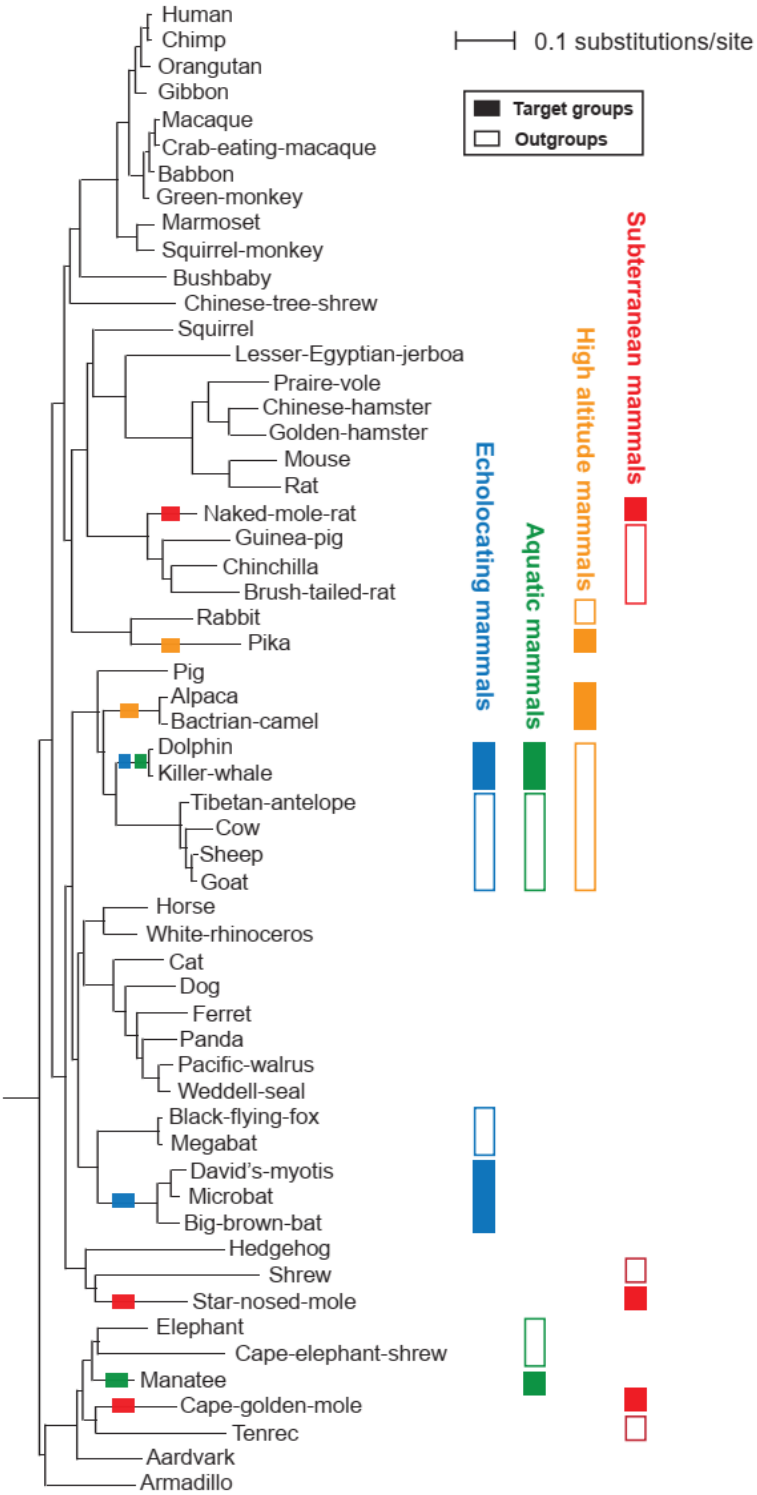
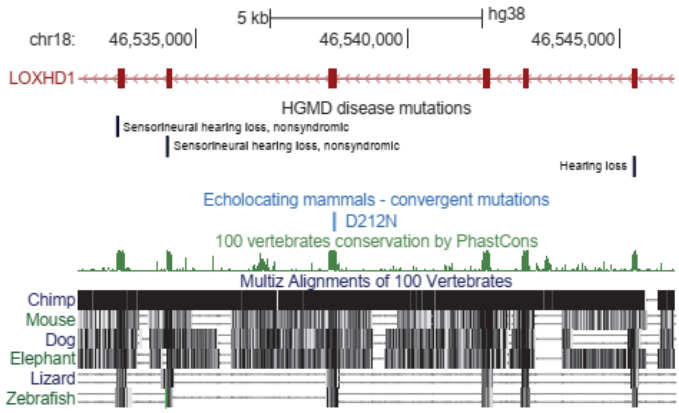


Figure S1. Weighted phylogenetic tree. A substitution per site weighted tree of the 57 species used in our screen. Filled rectangles mark target groups and empty rectangles mark outgroups for lineages with independent phenotypic evolution of echolocating (blue), aquatic (green), high-altitude (orange) and subterranean (red) lifestyles.



Human	I	N	A	D	T	G	A
Chimp	I	N	A	D	T	G	A
Gorilla	I	N	A	D	T	G	A
Orangutan	I	N	A	D	T	G	A
Gibbon	I	N	A	D	T	G	A
Rhesus	I	N	A	D	T	G	A
Crab-eating_macaque	I	N	A	D	T	G	A
Baboon	I	N	A	D	T	G	A
Green_monkey	I	N	A	D	T	G	A
Marmoset	I	N	A	D	T	G	A
Bushbaby	I	N	A	D	T	G	A
Chinese_tree_shrew	I	N	A	D	T	G	A
Squirrel	I	N	A	D	T	G	A
Lesser_Egyptian_jerboa	I	N	S	D	T	G	A
Prairie_vole	I	N	S	D	T	G	A
Golden_hamster	I	N	A	D	T	G	A
Mouse	I	N	A	D	T	G	A
Rat	V	N	A	D	T	G	A
Naked_mole-rat	I	N	A	D	T	G	A
Guinea_pig	I	N	A	D	T	G	A
Chinchilla	I	N	A	D	T	G	A
Brush-tailed_rat	I	N	S	D	T	G	A
Rabbit	I	N	A	D	T	G	A
Pika	I	N	A	D	T	G	A
Pig	I	N	S	D	T	G	A
Alpaca	I	N	A	D	T	G	A
Bactrian_camel	I	N	A	D	T	G	A
Dolphin	I	N	A	N	T	G	A
Killer_whale	I	N	A	N	T	G	A
Tibetan_antelope	I	N	A	D	T	G	A
Cow	I	N	A	D	T	G	A
Sheep	I	N	A	D	T	G	A
Domestic_goat	I	N	A	D	T	G	A
Horse	I	N	A	D	T	G	A
White_rhinoceros	I	N	A	D	T	G	A
Cat	I	N	A	D	T	G	A
Dog	I	N	A	D	T	G	A
Ferret	I	N	A	D	T	G	A
Panda	I	N	A	D	T	G	A
Pacific_walrus	I	N	A	D	T	G	A
Weddell_seal	I	N	A	D	T	G	A
Black_flying-fox	I	N	A	D	T	G	A
Megabat	I	N	A	D	T	G	A
David's_myotis_(bat)	V	N	A	N	T	G	A
Microbat	V	N	A	N	T	G	A
Big_brown_bat	V	S	A	N	T	G	A
Hedgehog	I	N	A	D	T	G	A
Shrew	I	N	A	D	T	G	A
Star-nosed_mole	I	N	A	D	T	G	A
Elephant	I	N	A	D	T	G	A
Cape_elephant_shrew	I	N	A	D	T	G	A
Manatee	I	N	A	D	T	G	A
Cape_golden_mole	I	N	A	D	T	G	A
Tenrec	I	N	A	D	T	G	A
Aardvark	I	N	A	D	T	G	A
Armadillo	I	N	A	D	T	G	A
Opossum	I	S	A	D	T	G	A
Tasmanian_devil	I	S	A	D	T	G	A
Platypus	V	D	A	D	T	G	A
Saker_falcon	V	D	A	D	T	G	A
White-throated_sparrow	V	D	A	D	T	G	A
Parrot	V	D	A	D	T	G	A
Rock_pigeon	V	D	A	D	T	G	A
Chicken	V	D	A	D	T	G	A
Green_seaturtle	V	D	A	D	T	G	A
Lizard	V	D	A	D	T	G	A
X_tropicalis	V	D	A	D	T	G	A
Coelacanth	V	D	A	D	T	G	A
Tetraodon	V	D	A	D	T	G	A
Fugu	V	D	A	D	T	G	A
Medaka	V	D	A	D	T	G	A
Stickleback	V	D	A	D	T	G	A
Zebrafish	V	N	A	D	T	G	A
Spotted_gar	V	D	A	D	T	G	A

Figure S2. Example of striking convergent substitution in echolocating mammals across a vertebrate multiple alignment. The functional enrichments reported in this study are composed of highly compelling molecular events. Convergent substitution D212N in *LOXHD1* gene is shown where the canonical amino acid is deeply conserved, suggesting that the independent D->N change in the two echolocating lineages is functionally important. *LOXHD1* is a well-established hearing loss gene in both humans and mice (1), but it has not been previously implicated in echolocation. While one could sift for this particular substitution in the hundreds of events we detect (Table S3), our test highlights it as part of a group of cochlear ganglion development genes (Fig. 2a, Tables 1 and 2), most affected by similarly striking convergent amino acid substitutions.

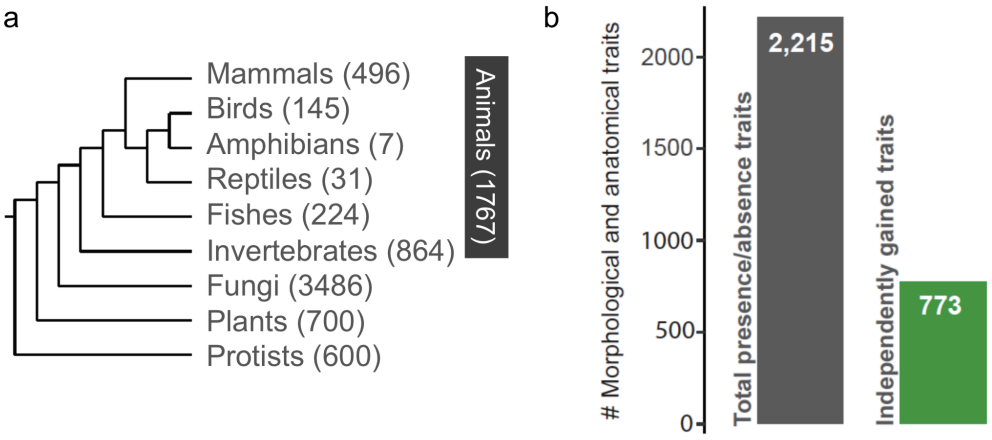


Figure S3. Available genotypic and phenotypic data for future exploration of the molecular basis of convergent evolution. **a,** The growing number of publicly available genomes at NCBI, by taxonomy. **b,** summary analysis of a large scale phenotypic study(2) that scores thousands of morphological and physiological traits (grey bar) for presence/absence reveals that a full 35% of scored traits (such as echolocation) are independently gained traits (green bar), conceptually amenable to our approach.

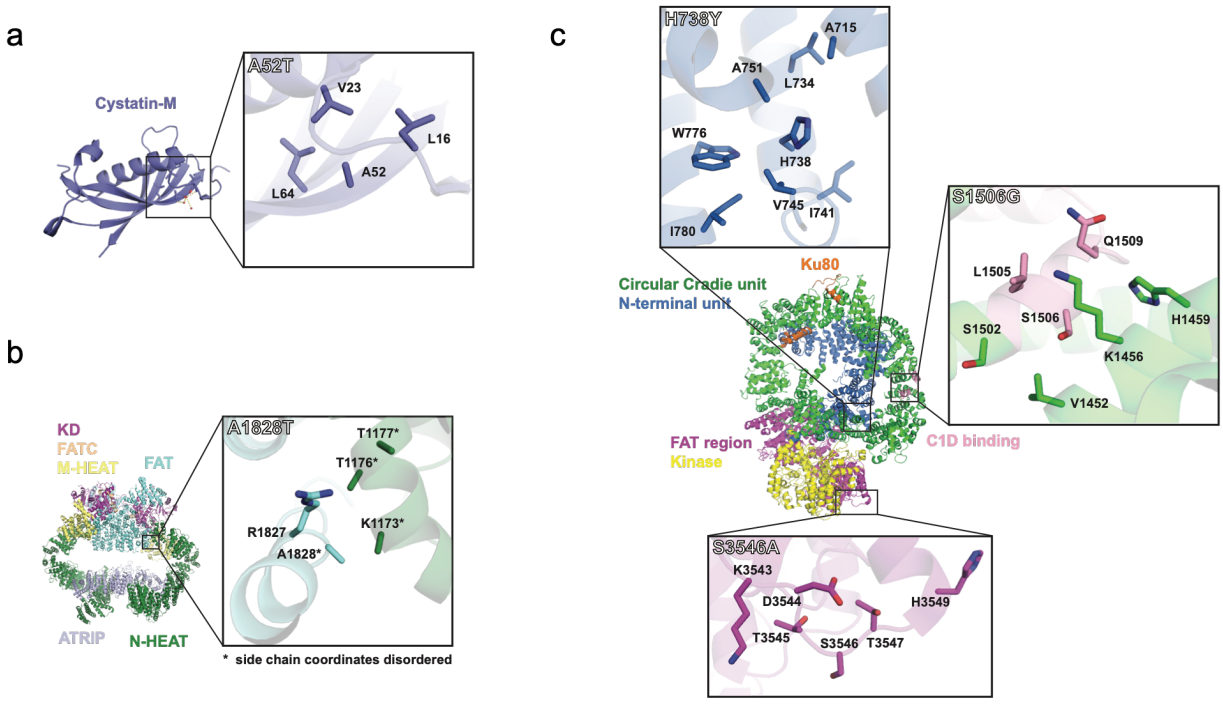


Figure S4. Protein structure evidence for functional significance of convergent substitutions. **a,** Based on a structural model of Cystatin-M (encoded by *CST6*) from PDB 4N6L, the side chain of residue A52 interacts with nearby hydrophobic residues in the wildtype protein. The convergent substitution at position 52 to a more polar threonine residue would thus destabilize the protein by disrupting the hydrophobic interaction. **b,** Structure of ATR-ATRIP complex from PDB 5YZ0. Residue A1828 is surrounded by hydrophilic residues (K1173, T1176, and T1177 of the N-HEAT domain and R1827 of the FAT domain). A1828T could enhance polar contacts with nearby residues and thus stabilize the binding interface between N-HEAT and FAT domains. **c,** Structure of DNA-PKcs in complex with Ku80 from PDB 5LUQ. Residue H738 is buried within the hydrophobic interior of DNA-PKcs. H738Y could stabilize the hydrophobic interaction and increase DNA-PKcs activity.

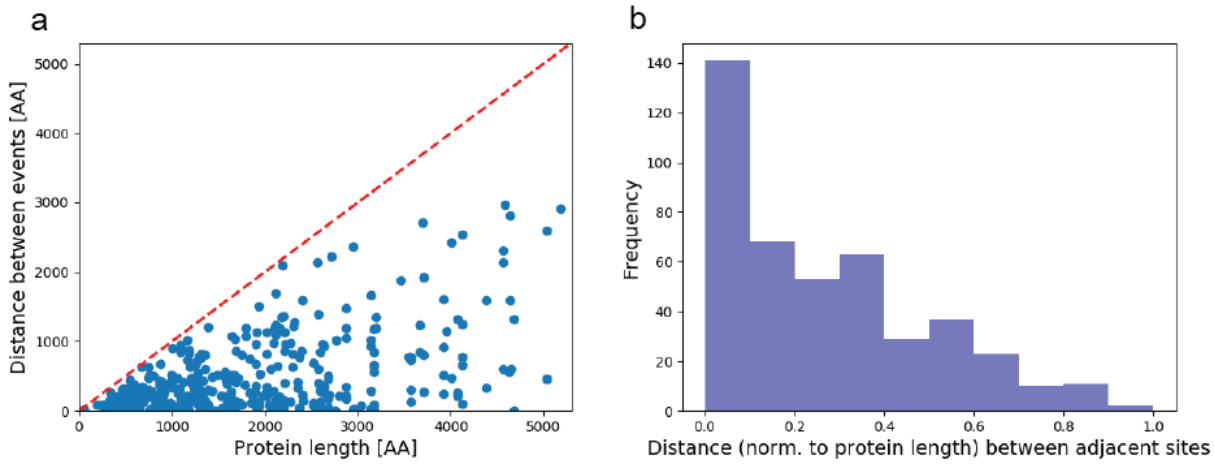


Figure S5. Pairwise distances between adjacent convergent residues within the same gene. a, Scatter plot of distance (in codon space) between adjacent amino acid (AA) convergence events in the same gene (Y-axis) and the corresponding protein length (X-axis) for all 437 pairs of same-gene adjacent convergent residues across the three convergence tests (echolocating, high-altitude and aquatic mammals). **b,** Histogram of the distance (in codon space, normalized to protein length) between same-gene adjacent convergent sites for all three convergence tests (echolocating, high-altitude, and aquatic mammals).

Species	Name	Assembly	Species	Name	Assembly
Panda	*Ailuropoda melanoleuca*	ailMel1	Ferret	*Mustela putorius furo*	musFur1
Cow	*Bos taurus*	bosTau8	David's myotis (bat)	*Myotis davidii*	myoDav1
Marmoset	*Callithrix jacchus*	calJac3	Microbat	*Myotis lucifugus*	myoLuc2
Bactrian camel	*Camelus ferus*	camFer1	Gibbon	*Nomascus leucogenys*	nomLeu3
Dog	*Canis lupus familiaris*	canFam3	Pika	*Ochotona princeps*	ochPri3
Domestic goat	*Capra hircus*	capHir1	Brush-tailed rat	*Octodon degus*	octDeg1
Guinea Pig	*Cavia porcellus*	cavPor3	Pacific walrus	*Odobenus rosmarus divergens*	odoRosDiv1
White Rhino	*Ceratotherium simum*	cerSim1	Killer whale	*Orcinus orca*	orcOrc1
Chinchilla	*Chinchilla lanigera*	chiLan1	Aardvark	*Orycteropus afer*	oryAfe1
Green monkey	*Chlorocebus sabaeus*	chlSab2	Rabbit	*Oryctolagus cuniculus*	oryCun2
Cape golden mole	*Chrysochloris asiatica*	chrAsi1	Bushbaby	*Otolemur garnettii*	otoGar3
Star-nosed mole	*Condylura cristata*	conCri1	Sheep	*Ovis aries*	oviAri3
Chinese Hamster	*Cricetulus griseus*	criGri1	Tibetan antelope	*Pantholops hodgsonii*	panHod1
Armadillo	*Dasypus novemcinctus*	dasNov3	Chimp	*Pan troglodytes*	panTro5
Tenrec	*Echinops telfairi*	echTel2	Baboon	*Papio hamadryas*	papAnu2
Cape elephant shrew	*Elephantulus edwardii*	eleEdw1	Orangutan	*Pongo pygmaeus abelii*	ponAbe2
Big brown bat	*Eptesicus fuscus*	eptFus1	Black flying fox	*Pteropus alecto*	pteAle1
Horse	*Equus caballus*	equCab2	Megabat	*Pteropus vampyrus*	pteVam1
Hedgehog	*Erinaceus europaeus*	eriEur2	Rhesus macaque	*Macaca mulatta*	rheMac8
Cat	*Felis catus*	felCat8	Rat	*Rattus norvegicus*	rn6
Naked mole rat	*Heterocephalus glaber*	hetGla2	Squirrel monkey	*Saimiri boliviensis*	saiBol1
Human	*Homo sapiens*	hg38	Shrew	*Sorex araneus*	sorAra2
Lesser Egyptian jerboa	*Jaculus jaculus*	jacJac1	Squirrel	*Spermophilus tridecemlineatus*	speTri2
Weddell seal	*Leptonychotes weddellii*	lepWed1	Pig	*Sus scrofa*	susScr3
Elephant	*Loxodonta africana*	loxAfr3	Manatee	*Trichechus manatus latirostris*	triMan1
Crab-eating macaque	*Macaca fascicularis*	macFas5	Chinese tree shrew	*Tupaia chinensis*	tupChi1
Golden hamster	*Mesocricetus auratus*	mesAur1	Bottlenose dolphin	*Tursiops truncatus*	turTru2
Prairie vole	*Microtus ochrogaster*	micOch1	Alpaca	*Vicugna pacos*	vicPac2
Mouse	*Mus musculus*	mm10			

Table S1. Species list. The 57 placental mammal species and genome assemblies used in this study.

Experiment	MGI terms tested	MGI-annotated sites	MGI-annotated genes	Convergent sites	Convergent genes	Divergent sites	Divergent genes
Echolocation	4,300	3,009,534	6,718	781	665	1,140	913
Aquatic	4,313	3,027,118	6,756	831	695	954	768
High altitude	4,165	2,646,617	6,139	977	792	1,306	988
Subterranean (moles-1)	4,102	2,515,604	5,916	1,333	1,044	2,259	1,437
Subterranean (moles-2)	4,131	2,606,536	5,993	1,086	891	1,690	1,197
Subterranean (moles-3)	4,241	2,815,825	6,465	1,075	883	1,646	1,201

Table S2. Key statistics from the molecular convergence test. For each of the six experiments, we show the number of MGI terms tested for enrichment of convergence and divergence, the total number of testable amino acid positions (*MGI-annotated sites*) and the number of genes containing them (*MGI-annotated genes*), the number of convergent and divergent substitutions detected, and the number of genes harboring them. Moles-1: cape golden mole and star-nosed mole; moles-2: naked mole rat and star-nosed mole; moles-3: cape golden mole and naked mole rat.

Convergent adaptation	Gene	Amino acid substitution	Codon coordinates (GRCh38/hg38)	Human disease variant (GRCh38/hg38)[I]	Source	Position relative to variant (in codon space) [II]
Echolocation	GJB2	F115Y	chr13:20189237-20189239	Sensorineural hearing loss (c.20189236delinsTA)	HGMD	+1
				Sensorineural hearing loss (c.20189238A>C)	HGMD	0
				Deafness (c.20189241T>C)	HGMD	-1
				Deafness, autosomal recessive 1 (c.20189242T>C)	HGMD	-1
	CDH23	N924S	chr10:71812589-71812591	Usher syndrome 1d (c.71812588G>A)	HGMD	-1
		D820E	chr10:71811415-71811417	Non-syndromic autosomal recessive deafness (c.71811412C>A)	HGMD	-1
	TMPRSS3	L385V	chr21:42376580-42376582	Deafness, non-syndromic, autosomal recessive (c.42376579A>G)	HGMD	+1
				Deafness, autosomal recessive 8,not specified (c.42376583A>G)	ClinVar	-1
Aquatic	LDLR	L422V	chr19:11113355-11113357	Hypercholesterolaemia (c.11113356T>C or c.11113356T>G	HGMD, ClinVar	0
	TRPV3	R148Q	chr17:3543496-3543498	Palmoplantar keratoderma, mutilating, with periorificial keratotic plaques (c.3543498C>G)	ClinVar	0
	XPA	V234L	chr9:97675559-97675561	Improved adduct removal (c.97675561C>A)	HGMD	0
High altitude	NPC1	I663M	chr18:23544485-23544487	Niemann-Pick disease C (c.23544484C>T)	HGMD	+1
		F356L	chr18:23556501-23556503	Niemann-Pick disease C (NM_000271.4(NPC1):c.1065delC (p.Phe356Serfs))	ClinVar	0

[I] Human disease-modulating variants in HGVS nomenclature

[III] Position of convergent substitution relative to human disease-modulating variant: 0=exact overlap; +1=convergent substitution is one codon upstream of variant (i.e. towards N-terminus); -1=convergent substitution is one codon downstream of variant (i.e. towards C-terminus)

Table S3. Convergent sites proximal to known human disease-modulating variants. Shown are the parallel convergent sites from Table 1 that overlap or are within one codon of a known human disease variant from HGMD(3) or ClinVar(4).

Convergent adaptation	Term ID	Term	Total # term genes	Total # term amino acid positions	# convergent term genes	# convergent term positions	# divergent term genes	# divergent term positions	Fold enrichment (convergent)	Fold enrichment (divergent)	p-value (convergent)	p-value (divergent)	q-value (convergent)	q-value (divergent)
Echolocation	MP:0002857	cochlear ganglion degeneration	53	30304	18	25	11	15	3.2	1.3	7.19E-07	1.82E-01	2.17E-03	1
	MP:0002855	abnormal cochlear ganglion morphology	82	41563	21	28	13	18	2.6	1.1	7.34E-06	3.16E-01	1.05E-02	1
Aquatic	MP:0001192	scaly skin	49	24105	15	27	10	12	4.1	1.6	1.79E-09	8.42E-02	3.86E-06	1
	MP:0001240	abnormal epidermis stratum corneum morphology	125	54056	23	42	22	32	2.8	1.9	3.90E-09	6.84E-04	5.60E-06	0.109
	MP:0001198	tight skin	17	14592	4	17	5	8	4.2	1.7	1.04E-06	9.44E-02	6.42E-04	1
	MP:0001242	hyperkeratosis	90	35761	16	28	15	21	2.9	1.9	1.31E-06	5.76E-03	7.05E-04	0.421
	MP:0010955	abnormal respiratory electron transport chain	55	16684	16	18	3	3	3.9	0.6	1.48E-06	8.96E-01	7.10E-04	1
	MP:0002841	impaired skeletal muscle contractility	35	30160	13	25	8	14	3	1.5	1.79E-06	1.01E-01	7.73E-04	1
	MP:0006036	abnormal mitochondrial physiology	115	46768	26	32	5	5	2.5	0.3	3.88E-06	9.99E-01	1.46E-03	1
	MP:0001212	skin lesions	111	53776	23	35	22	28	2.4	1.7	4.07E-06	7.95E-03	1.46E-03	0.475
	MP:0010954	abnormal cellular respiration	97	33894	21	25	5	6	2.7	0.6	1.28E-05	9.56E-01	3.96E-03	1
	MP:0001231	abnormal epidermis stratum basale morphology	55	34579	12	24	10	14	2.5	1.3	4.88E-05	2.08E-01	1.00E-02	1
	MP:0010814	absent alveolar lamellar bodies	11	5997	2	9	2	3	5.5	1.6	5.45E-05	2.94E-01	1.02E-02	1
	MP:0004091	abnormal Z lines	31	24304	8	19	5	8	2.8	1	6.46E-05	4.99E-01	1.11E-02	1
	MP:0003941	abnormal skin development	22	14812	5	14	3	3	3.4	0.6	8.52E-05	8.45E-01	1.36E-02	1
	MP:0002796	impaired skin barrier function	64	31323	9	22	11	15	2.6	1.5	8.37E-05	7.58E-02	1.36E-02	1
	MP:0005048	thrombosis	91	48149	22	28	12	19	2.1	1.3	2.32E-04	1.91E-01	2.99E-02	1
	MP:0001429	dehydration	77	38747	13	24	7	10	2.3	0.8	2.61E-04	7.77E-01	3.22E-02	1
	MP:0000764	abnormal tongue epithelium morphology	21	16673	6	14	3	5	3.1	1	2.82E-04	6.04E-01	3.38E-02	1
	MP:0010813	abnormal alveolar lamellar body morphology	28	13026	5	12	5	6	3.4	1.5	3.35E-04	2.31E-01	3.90E-02	1
	MP:0011159	abnormal epidermal-dermal junction morphology	10	13145	6	12	2	4	3.3	1	3.63E-04	5.94E-01	4.12E-02	1
	MP:0001209	spontaneous skin ulceration	32	15456	10	13	6	6	3.1	1.2	4.52E-04	3.61E-01	4.75E-02	1
	MP:0000751	myopathy	44	35397	10	22	11	20	2.3	1.8	4.42E-04	1.02E-02	4.75E-02	0.529
High altitude	MP:0005629	abnormal lung weight	52	25252	16	25	5	10	2.7	0.8	1.35E-05	7.97E-01	1.95E-02	1
	MP:0005350	increased susceptibility to autoimmune disorder	126	38001	18	32	16	21	2.3	1.1	2.25E-05	3.31E-01	1.95E-02	1
	MP:0002790	decreased circulating follicle stimulating hormone level	26	10872	7	14	3	10	3.5	1.9	7.50E-05	4.67E-02	2.81E-02	0.431
	MP:0004953	decreased spleen weight	65	27766	17	25	14	22	2.4	1.6	6.14E-05	2.29E-02	2.81E-02	0.294
	MP:0005027	increased susceptibility to parasitic infection	75	17160	11	18	12	14	2.8	1.7	1.03E-04	4.95E-02	3.31E-02	0.445

Table S4. All ontology terms passing the criteria for the combined convergence-divergence test (see main text). For each experiment, the terms are sorted by *q-value* (convergent) and then fold enrichment (convergent). The top term from each experiment, discussed in the main text, is highlighted in blue.

References for Supplementary Information

1. Grillet N, et al. (2009) Mutations in LOXHD1, an evolutionarily conserved stereociliary protein, disrupt hair cell function in mice and cause progressive hearing loss in humans. *Am J Hum Genet* 85(3):328–337.

2. O'Leary MA, et al. (2013) The Placental Mammal Ancestor and the Post–K-Pg Radiation of Placentals. *Science* 339(6120):662–667.

3. Stenson PD, et al. (2003) Human Gene Mutation Database (HGMD): 2003 update. *Hum Mutat* 21(6):577–581.

4. Landrum MJ, et al. (2018) ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res* 46(D1):D1062–D1067.