Supplementary Table 1 A randomly selected subset of 100 OMIM record pairs cross-referenced by the OMIM curators. To be sure the 7,000 cross referenced OMIM records in fact were pairs of highly overlapping phenotypes, we manually cuated a random subset of 100 record pairs. This investigation shows that 94 of the record pairs had a high degree of phenotypic overlap. 94 our of 100 record pairs are true positives (TP) (i.e., record pairs cross-referenced because of a high degree of phenotypic overlap. 6 record pairs were false positives (FP) (i.e., record pairs that were cross-referenced for other reasons than a high degree of phenotypic overlap). Based on this investigation we concluded that the full set of 7,000 pairs could be used as a benchmarking reference for the phenotype similarity score.

In the table we have included a comment stating why there is a phenotypic overlap. If the OMIM text regarding the overlap is very clear, or the overlap is obvious, we have not included a reference to Pubmed, but simply stated OMIM as the reference. If the overlap is not obvious, or the statements in OMIM were ambiguous, we referenced one or more articles describing the phenotypic overlap.

Disease (OMIM ID)	Disease (OMIN ID)	Comment	Pubmed ID
Walker-Warburg syndrome Muscular dystrophy, congenital, mental retardation included #236670	FUKUYAMA CONGENITAL MUSCULAR DYSTROPHY #253800	TP: Two different forms and loci of Muscular dystrophy and mental retardaion	9763366
Muir-Torre syndrome #158320	Colorectal cancer, hereditary nonpolyposis #120435	TP: Muir-Torre syndrome is a subtype of Colorectal cancer, hereditary nonpolyposis	10544223
Bardet-Biedl syndrome #209900	Meckel syndrome, type 1 #249000	TP: cystic kidney dysplasia and polydactyly occurs in both syndromes	1488972
Achalasia-Addisonianism-Alacrima syndrome #231550	Addison disease and cerebral sclerosis #300100	TP: “The association of adrenal and neurologic disease in #231550 is similar to that in #300100	OMIM
Knobloch syndrome #267750	Knobloch syndrome, type II %608454	TP: Two different forms and loci of Knobloch syndrome	OMIM
Blood group, Froese #601551	Blood group, Diego system #110500	TP: Blood group Froese is part of the Diego system. Same molecular basis. “Antigen polymorphism is based on a change in the SLC4A1 gene”	OMIM 11061863
Granulomatous disease, cytochrome b-positive #233710	Granulomatous disease, cytochrome b-negative #233690	TP: Two different forms and loci of Granulomatous disease	OMIM
Glaucoma 1, open angle, A #137750	Glaucoma 1, open angle, C %601682	TP: Two different loci of Glaucoma 1	11436127
Reis-Bucklers Corneal	Thiel-Behnke Corneal	TP: Two different forms	7671605

dystrophy #608470	dystrophy %602082	and loci of Corneal dystrophy	
Hypophosphatemic rickets #300554	Dent disease 1 (Nephrolithiasis, hypercalciuria) #300009	TP: “Hypophosphatemic rickets is a form of hypercalciuric nephrolithiasis”, “Dent disease complex”	15558518 9452994
Mullerian Aplasia #158330	Maturity-onset diabetes of the young 1 #125850	TP: Mullerian Aplasia like MODY1 is characterized by frequent complications of diabetic nature including effects on the kidneys. Both conditions caused by mutations in transcription factors influencing e.g. insulin transcription	10484768
Situs invertus viscerum #270100	Primary ciliary dyskinesia #242650	TP: Two associated diseases making up Kartagener syndrome #244400	12142464
Hypertension with brachydactyly %112410	Brachydactyly, type E #113300	TP: Two forms and loci of brachydactyly and hypertension	OMIM 9696728
Primary lateral sclerosis #606353	Spastic paralysis #607225	TP: “lateral sclerosis and spastic paralysis are allelic disorders with overlapping phenotypes”	OMIM
Cataract, nonnuclear congenital %601286	Cataract, embryonic nuclear #604307	TP: Two different forms but associated with same loci	OMIM 8733140
Epidermolysis bullosa dystrophica #132000	Epidermolysis bullosa simplex #131900	TP: Two different forms and loci of Epidermolysis bullosa	OMIM 7682883
Rett syndrome #312750	Incontinentia pigmenti #308300	TP: Some times Rett syndrome co-occurs with incontinentia pigmenti	OMIM 10577905
FG syndrome including partial agenesis of corpus callosum %305450	Corpus callosum, partial agenesis #304100	TP: Phenotypically overlapping diseases	OMIM
Medullary cystic kidney disease 1 %174000	Polycystic kidney disease #173900	TP: Two different form and loci of cystic kidney disease	OMIM 11261687
Nanophthalmos 1 %600165	Nanophthalmos 2 #609549	TP: Two different loci of Nanophthalmos	OMIM
Spondyloepimetaphyseal dysplasia with multiple dislocations %603546	Spondyloepimetaphyseal dysplasia, sponastrime type %271510	TP: Two different forms and loci of Spondyloepimetaphyseal dysplasia	9678701
Noncompaction of left	Noncompaction of left	TP: Two different loci of	15173023

ventricular myocardium, 1 #604169	ventricular myocardium, 2 %609470	Noncompaction of left ventricular myocardium	
Periodic paralysis, potassium-sensitive #170390	Hypokalemic periodic paralysis #170400	TP: Two different forms and loci of periodic paralysis	12536108
Craniosynostosis, midfacial hypoplasia and foot abnormalies #123150	Apert syndrome #101200	TP: Two different forms and loci of craniosynostosis	11121055
Preeclampsia/eclampsia 3 %609403	Preeclampsia/eclampsia 1 %189800	TP: Two different forms and loci of Preeclampsia/eclampsia	OMIM
Gastric cancer #137215	Helicobacter pylori infection #600263	TP: “Helicobacter pylori infection is an established risk factor of gastric cancer, however gastric cancer occurs in only a very small proportion of people infected with the organism”	12115538
Colorectal cancer #120435	Endometrial cancer #608089	TP: Endometrial cancer is often co-occurring with colorectal cancer	OMIM
Moyamoya disease 3 %608796	Moyamoya disease 1 %252350	TP: Two different loci of Moyamoya disease	OMIM
Charcot-Marie-Tooth disease 6 #601152	Charcot-Marie-Tooth disease 1B #118200	TP: Two different forms and loci of Charcot-Marie-Tooth disease	OMIM
Premature ovarian failure 1 %311360	Addison disease %240200	TP: Premature ovarian failure occasionally occur in patients with Addison disease	OMIM 2686250
Pseudohypoaldosteronism, type 1 #264350	Pseudohypoaldosteronism, type II #145260	TP: Two different forms and loci of Pseudohypoaldosteronism	2137831
Specific granule deficiency #245480	Chediak-Higashi syndrome #214500	TP: Both disorders characterized by severe defensin deficiency	2841356
Osteosarcoma #259500	Chondrosarcoma #215300	TP: Two different forms and loci of sarcoma	OMIM
Ichthyosis, lamellar 1 #242300	Ichthyosis, lamellar 2 #601277	TP: Two different forms and loci of Ichthyosis	OMIM
von Hippel-Lindau syndrome #193300	Pheochromocytoma #171300	TP: “Pheochromocytomas most commonly occur as part of several syndromes including von Hippel-Lindau syndrome”	OMIM
Arthrogryposis, type 5 %108145	Arthrogryposis, type 2B #601680	TP: Two different forms and loci of Arthrogryposis	OMIM 8923935
Dementia with spastic ataxia #176500	Creutzfeldt-Jakob disease #123400	TP: Overlapping phenotypes. Dementia and another form of ataxia	OMIM

		is part of the clinical synopsis for Creutzfeldt-Jakob disease.	
Coffin-Siris syndrome %135900	Biotinidase deficiency #253260	TP: Case report identifying an association between the two diseases	2373113
Systemic lupus erythematosus 7 %610065	Systemic lupus erythematosus #152700	TP: Different loci of Systemic lupus erythematosus	OMIM 16642431
Pancreatic carcinoma #260350	Peutz-Jeghers syndrome #175200	TP: Increased risk of pancreatic cancer in patients with Peutz-Jeghers syndrome	OMIM
Pseudovitamin D3 deficiency Rickets due to 25-hydroxylase deficiency #600081	Vitamin D-dependent Rickets type 1	TP: Two different forms and loci of Vitamin D-dependent Rickets	OMIM
Ceroid lipofuscinosis, neuronal 6 #601780	Ceroid lipofuscinosis, neuronal 3 #204200	TP: Two different forms and loci of Ceroid lipofuscinosis, neuronal	OMIM 9097964
Thyrotoxic periodic paralysis #188580	Hashimoto thyroiditis %140300	TP: Two phenotypically overlapping diseases	OMIM 3839536
Pilomatrixoma #132600	Rubinstein-Taybi syndrome #180849	TP: “Occurrence of pilomatrixomas in Rubinstein-Taybi syndrome”	9557902
Myopathy, myofibrillar, desmin-related #601419	Myopathy, myofibrillar, myotilin-related #609200	TP: Two different forms and loci of Myopathy	OMIM
Hyperthyroidism #609152	Hypothyroidism #275200	FP: Both conditions can be caused by mutations in TSHR, but clinical phenotypes have little overlap	OMIM
Charcot-Marie-Tooth disease dominant intermediate D #607791	Charcot-Marie-Tooth disease dominant intermediate A %606483	TP: Two different forms and loci of Charcot-Marie-Tooth disease	OMIM
Glioma of brain #137800	Cowden disease #158350	FP: From abstract, “the patient did not show any of the clinical signs of Cowden disease”.	12085208
Congenital alopecia #601705	Omenn syndrome #603554	TP: Two sisters with congenital alopecia #601705. Alopecia is highly frequent in Omenn syndrome	11795679
Hypomagnesemia 2 #154020	Hypomagnesemia with secondary hypocalcemia #602014	TP: Two different forms and loci of Hypomagnesemia	9915957
Craniosynostosis, Type 1	Greig	TP: Craniosynostosis is	OMIM

%123100	cephalopolysyndactyly syndrome #175700	part of the clinical synopsis of Greig cephalopolysyndactyly, although craniosynostosis is described three times in the literature as NOT being part of Greig cephalopolysyndactyly 7521123, 8188211, 223435	
Chondrodysplasia punctata 2 #302960	Epilepsy, with mental retardation. %300088	TP: Several reports where Chondrodysplasia punctata and epilepsy/mental retardations are connected	15658616
Pancreatic insufficiency and bone marrow dysfunction #260400	Asphyxiating thoracic dystrophy %208500	TP: Skeletal abnormalities part of both diseases	OMIM 999326
Muscular dystrophy, Type 1 F %608423	Muscular dystrophy, Type 1 A #159000	TP: Two different forms and loci of Muscular dystrophy	11222786
Attention deficit-hyperactivity disorder 2 %608904	Attention deficit-hyperactivity disorder 1 %608903	TP: Two different loci of Attention deficit-hyperactivity disorder	15297934
Albright hereditary osteodystrophy #103580	Major affective disorder %309200	FP: The link of the diseases are due to their mode of inheritance	6770678
Major affective disorder 1 %125480	Major affective disorder 2 %309200	TP: Two different loci of Major affective disorder	OMIM 11149935
Malignant hyperthermia 1 #145600	Malignant hyperthermia 4 %600467	TP: Two different loci of Malignant hyperthermia	OMIM 12522565
Cataract, anterior polar 2 %601202	Cataract, anterior polar 1 %115650	TP: Two different loci of anterior polar cataract	OMIM 8852669
Prostate cancer aggressiveness QTL %607592	Prostate cancer #176807	TP: Two different loci of Prostate cancer	10825281
Parkinson disease 9 %606693	Pallidopyramidal syndrome %260300	TP: “clinical features similar to those of idiopathic Parkinson disease and Pallidopyramidal syndrome”	8085432
Hypertension #145500	Pseudohypoaldosteronism 1 #264350	FP: Opposite clinical phenotypes. However, both conditions are caused by changed renal tubular Na+ absorption (hypertension: increased. PHA-1: decreased)	15931322

Spinocerebellar ataxia 2 #183090	Spinocerebellar ataxia 6 #183086	TP: Two different forms and loci of Spinocerebellar ataxia	16613893
Amyloidosis corneal #204870	Amyloidosis V #105120	TP: Two forms and loci of amyloidosis	OMIM 10729296
Globoid cell leukodystrophy #245200	Metachromatic leukodystrophy #250100	TP: Two forms and loci of leukodystrophy	16763970
Granunomatous disease, x-linked #306400	Duchenne muscular dystrophy #310200	FP: Case study. One patient had 3 x-linked disorders incl. Granunomatous disease and Duchenne muscular dystrophy. Deletion of part of Xp21	4039107
Lymphedema-Distichiasis syndrome #153400	Lymphedema, hereditary II #153200	TP: “Allelic disorders with overlapping phenotypes include Lymphedema type II”	10086462
Celiac disease susceptibility to 1 #212750	Celiac disease susceptibility to 3 #609755	TP: Two different loci of Celiac disease	16820991
GM2-gangliosidosis type II #268800	Gaucher disease type 1 #230800	TP: “Gaucher disease, a related disorder that involves glycosphingolipid storage in peripheral tissues but not in the central nervous system”	OMIM 16854371
Inflammatory bowel disease 2 %601458	Inflammatory bowel disease 1 #266600	TP: Two different loci of Inflammatory bowel disease	16873002
Episodic ataxia, type 1 #160120	Episodic ataxia, type 2 #108500	TP: Two different forms and loci of Episodic ataxia	9390841
Brachydactyly, type A1 #112500	Brachydactyly, type A1, B %607004	TP: Two different loci of Brachydactyly, type A1	11897820
Usher syndrome, type IIB %276905	Usher syndrome, type IIA #276901	TP: Two different forms and loci of Usher syndrome type II	11950859
Craniofacial-deafness-hand syndrome #122880	Waardenburg syndrome, type 1 #193500	TP: Several overlapping clinical traits between the diseases.	14556253
Major depressive disorder #608516	Major depressive disorder 2 %608691	TP: Two different loci of Major depressive disorder	OMIM 8966556
Waardenburg syndrome type IIA #193510	Waardenburg syndrome type IIB %600193	TP: Two different forms and loci of Waardenburg syndrome	331943
Orofacial cleft 5 #608874	Orofacial cleft 1 %119530	TP: Two different loci of Orofacial cleft	16282779
Thanatophoric dysplasia #187600	Asphyxiating thoracic dystrophy %208500	TP: Many phenotypically overlapping features	OMIM
Night Blindness, congenital	Aland Island eye disease	TP: “congenital stationary	7612552

type 1 #310500	%300600	night blindness and Aland Island eye disease could be considered as a single entity”	
Charcot-Marie-Tooth disease type 2A1 #118210	Charcot-Marie-Tooth disease type 2A2 #609260	TP: Two different forms and loci of Charcot-Marie-Tooth disease	OMIM
Myopia 12 %609995	Myopia 2 %160700	TP: Two different forms and loci of Myopia	OMIM
Migraine, familial typical, 2 %300125	Migraine with or without aura, 1 %157300	TP: Two different forms and loci of Migraine	OMIM
Dyslexia, susceptibility to 1 #127700	Dyslexia, susceptibility to 8 %608995	TP: Two different forms and loci of Dyslexia	OMIM
Chordoma %215400	Tuberous sclerosis complex #191100	TP: “identified 3 reports of chordomas found in patients with tuberous sclerosis complex”	15236319
Granulomatous disease, cronic, x-linked #306400	Granulomatous disease, cronic, autosomal recessive #233710	TP: Two forms and loci of Granulomatous disease	OMIM
Ulcerative colitis #191390	Inflammatory bowel disease 2 %601458	TP: Two different forms and loci of inflammatory bowel disease	OMIM
Velocardiofacial syndrome #192430	Asperger syndrome 1 %608638	TP: VCF is caused by mutation in TBX1. There is evidence for that “TXB1 haploinsufficiens may cause psychiatric disorders like ASP-1, although based only on casuistic reports”	16684884
Hair color 3 %601800	Eye color 3 %227220	TP: Association study shows the two traits are significantly inherited together	8875191
Alopecia-mental retardation %203650	Hair-brain syndrome #234050	TP: Overlapping phenotypes; mental retardation and abnormal hair growth.	OMIM
Epilepsy, idiopathic generalized #600669	Epilepsy, grand mal seizures on awakening #607628	TP: Two different forms and loci of epilepsy	OMIM
Hereditary myopathy with early respiratory failure #603689	Duchenne muscular dystrophy #310200	TP: Two different forms and loci of myopathy	OMIM
Wolfram syndrome 1 #222300	Wolfram syndrome 2 %604928	TP: Two different loci of Wolfram syndrome	10739754
Pancreatic carcinoma #260350	Melanoma, cutaneous malignant %155600	TP: “Several familial cancer syndromes increase the risk of	OMIM

		pancreatic cancer…"	
Piebald trait #172800	Fanconi anemia #227650	TP: "White forelock and patches leukoderma occur also in Fanconi anemia"	OMIM
Glaucoma 1, open angle, A #137750	Glaucoma 1, open angle, J %608695	TP: Two different loci of Glaucoma 1, open angle	15108121
Oral-facial-digital syndrome IX %258865	Oral-facial-digital syndrome II %252100	TP: Two different forms and loci of Oral-facial-digital syndrome	8074150
Hydranencephaly, fowler type %225790	Fetal akinesia deformation sequence (FADS) %208150	TP: Case study, both diseases included	11857548
Vitamin D-dependent rickets type II %600785	Vitamin D-dependent rickets type I #264700	TP: Two different forms and loci of Vitamin D-dependent rickets	OMIM
Cholestasis, progressive familial intrahepatic 2 #601847	Cholestasis, progressive familial intrahepatic 1 #211600	TP: Two different forms and loci of Cholestasis, progressive familial intrahepatic	OMIM
Vohwinkel syndrome #124500	Vohwinkel syndrome with ichthyosis #604117	TP: Two different forms and loci of Vohwinkel syndrome	OMIM