The Discovery of a *LEMD2*-Associated Nuclear Envelopathy with Early Progeroid Appearance Suggests Advanced Applications for AI-Driven Facial Phenotyping

Felix Marbach,[1,2,19] Cecilie F. Rustad,[3,19] Angelika Riess,[4,19] Dejan Đukić,[5] Tzung-Chien Hsieh,[6] Itamar Jobani,[7] Trine Prescott,[8] Andrea Bevot,[9] Florian Erger,[1,2] Gunnar Houge,[10,11] Maria Redfors,[12,13] Janine Altmueller,[14] Tomasz Stokowy,[11] Christian Gilissen,[15] Christian Kubisch,[16] Emanuela Scarano,[17] Laura Mazzanti,[17] Torunn Fiskerstrand,[10,11,18] Peter M. Krawitz,[6] Davor Lessel,[16,20] and Christian Netzer[1,2,20,*]

Over a relatively short period of time, the clinical geneticist's "toolbox" has been expanded by machine-learning algorithms for image analysis, which can be applied to the task of syndrome identification on the basis of facial photographs, but these technologies harbor potential beyond the recognition of established phenotypes. Here, we comprehensively characterized two individuals with a hitherto unknown genetic disorder caused by the same *de novo* mutation in *LEMD2* (c.1436C>T;p.Ser479Phe), the gene which encodes the nuclear envelope protein LEM domain-containing protein 2 (LEMD2). Despite different ages and ethnic backgrounds, both individuals share a progeria-like facial phenotype and a distinct combination of physical and neurologic anomalies, such as growth retardation; hypoplastic jaws crowded with multiple supernumerary, yet unerupted, teeth; and cerebellar intention tremor. Immunofluorescence analyses of patient fibroblasts revealed mutation-induced disturbance of nuclear architecture, recapitulating previously published data in LEMD2-deficient cell lines, and additional experiments suggested mislocalization of mutant LEMD2 protein within the nuclear lamina. Computational analysis of facial features with two different deep neural networks showed phenotypic proximity to other nuclear envelopathies. One of the algorithms, when trained to recognize syndromic similarity (rather than specific syndromes) in an unsupervised approach, clustered both individuals closely together, providing hypothesis-free hints for a common genetic etiology. We show that a recurrent *de novo* mutation in *LEMD2* causes a nuclear envelopathy whose prognosis in adolescence is relatively good in comparison to that of classical Hutchinson-Gilford progeria syndrome, and we suggest that the application of artificial intelligence to the analysis of patient images can facilitate the discovery of new genetic disorders.

Nuclear envelopathies are rare disorders caused by mutations in genes encoding proteins of the nuclear envelope (NE); these mutations result in a variety of phenotypes, some of which mimic aspects of human aging.[1] A prototypical example is Hutchinson-Gilford progeria syndrome (HGPS, MIM: 176670), which is typically accompanied by early-onset atherosclerosis, leading to cardiovascular complications and mortality in the late second decade of life (see Web Resources).[2] It is caused by a recurrent *de novo* synonymous splicing mutation in *LMNA* (MIM: 150330), which encodes the splicing isoforms lamin A and C (A-type lamins), the main components of the nuclear lamina in somatic cells.[3] Many other proteins are "anchored" within the meshwork created by lamins, and some of the respective genes, such as *LEMD3* (also known as *MAN1* [MIM: 607844]), *BANF1* (MIM: 603811), and *EMD* (MIM: 300384), are associated with nuclear envelopathies (see Web Resources).[4,5] At the cellular level, mutations in these genes cause disruption of the nuclear architecture and gene regulation. The integral NE protein LEMD2 (LEM domain-containing protein 2) is functionally and physically linked to A-type lamins, and experimental data suggest it has a role in the structural organization of the NE[6] as well as chromatin binding and distribution.[7,8] Although a single *LEMD2* founder mutation has been implicated in an autosomal-recessive form of congenital cataracts (MIM: 212500),[9] *LEMD2* has not been associated with a complex syndromic phenotype to date.

[1]Faculty of Medicine, University of Cologne, Cologne, 50931, Germany; [2]Institute of Human Genetics, University Hospital Cologne, Cologne, 50931, Germany; [3]Section of Clinical Genetics, Department of Medical Genetics, Oslo University Hospital, Oslo, 0424, Norway; [4]Institute of Medical Genetics and Applied Genomics, University of Tuebingen, Tuebingen, 72076, Germany; [5]Gregor Mendel Institute, Vienna, 1030, Austria; [6]Institute for Genomic Statistics and Bioinformatics, University Hospital Bonn, Bonn, 53127, Germany; [7]Private contributor, Tel Aviv, 6523217, Israel; [8]Department of Medical Genetics, Telemark Hospital Trust, Skien, 3710, Norway; [9]Department of Paediatric Neurology and Developmental Medicine, University Children's Hospital Tuebingen, Tuebingen, 72076, Germany; [10]Department of Medical Genetics, Haukeland University Hospital, Bergen, 5021, Norway; [11]Department of Clinical Science, University of Bergen, Bergen, 5021, Norway; [12]Department of Maxillofacial Radiology, Faculty of Dentistry, University of Oslo, Oslo, 0440, Norway; [13]National Resource Centre for Oral Health in Rare Medical Conditions, Lovisenberg Diakonale Hospital, Oslo, 0440, Norway; [14]University of Cologne, Cologne Center for Genomics, Cologne, 50931, Germany; [15]Department of Human Genetics, Radboud Institute for Molecular Life Sciences, Radboud University Medical Center, Nijmegen, 6525 GA, the Netherlands; [16]Institute of Human Genetics, University Medical Center Hamburg-Eppendorf, Hamburg, 20246, Germany; [17]Rare Disease Unit, Pediatric Clinic, Medical and Surgical Science Department, S.Orsola-Malpighi Hospital, University of Bologna, Bologna, 40138, Italy
[18]Deceased
[19]These authors contributed equally
[20]These authors contributed equally
*Correspondence: christian.netzer@uk-koeln.de
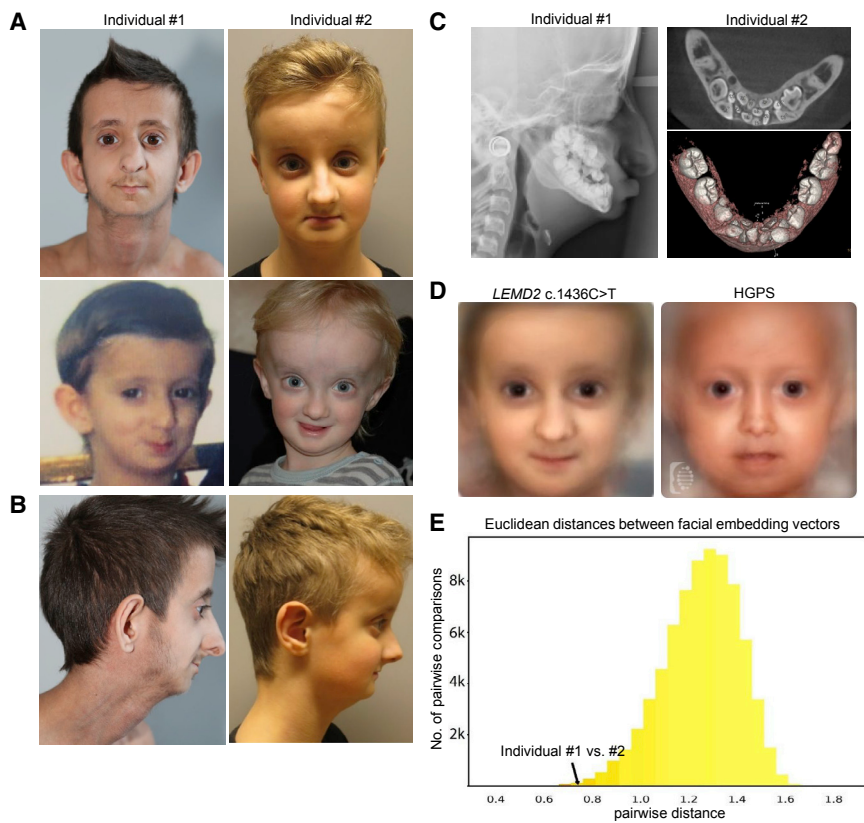https://doi.org/10.1016/j.ajhg.2019.02.021.

Figure 1. Clinical Phenotype and Image Analysis of Individuals 1 and 2

(A) Facial phenotypes of individuals 1 (left) and 2 (right). Individual 1 is shown at the ages of 16 years (upper panel) and 3 years (lower panel); individual 2 is shown at the ages of 10 years and 2 years (upper and lower panel, respectively). Note the protruding eyes; small mandibula and oral aperture; and triangular face of both individuals; this latter feature is more pronounced at a younger age.

(B) (Left to right) The profiles of individuals 1 and 2 as seen at the ages of 16 and 11 years, respectively.

(C) Dental features of individuals 1 and 2. Left panel: A cranial X-ray of individual 1 shows crowded, unerupted teeth in the upper and lower jaw at the age of 13 years. Right panels: Cranial CBCT scans of individual 2 revealed complete non-erupted primary dentition and agenesis of multiple teeth in the non-erupted permanent dentition in the mandible at the age of 9 years.

(D) Composite images for LEMD2 and HGPS were averaged from photos of both individuals (left) and HGPS patients (right), as described in the Supplemental Material and Methods, and manually retouched.

(E) Histogram of the pairwise distances among all cases of the cohort. Comparisons of the same disease entity were used for adding a red blend to the respective bins according to their proportion. At the left side of the distribution, where the two individuals with the LEMD2 mutation are also posed, the percentage of pairs with the same disease-causing gene increases.

We performed molecular analyses and artificial-intelligence-driven facial phenotyping on two individuals who are of different ethnicities and share similar facial features and physical anomalies. Both individuals and their families gave their written informed consent to participate in this study and to undergo facial analysis by specialized software (University of Cologne). Genetic analyses were approved by the ethics committees of the University of Cologne (individual 1) and the Regional Committee for Medical and Health Research Ethics in Western Norway (individual 2).

Individuals 1 and 2 presented at different university hospitals (Bologna, Italy and Oslo, Norway). Because of the individuals' early progeria-like appearance with little subcutaneous fat and triangular facies, clinicians suspected HGPS during the first years of life in both cases. However, the individuals' further development was encouraging, and the initial suspected diagnosis was discarded because neither developed joint contractures, alopecia, or clinical signs of atherosclerosis. Psychomotor development was normal, but poor linear growth was noted. Both had severely delayed dentition with crowded, mostly unerupted teeth in the upper and lower jaw (Figure 1C). Oral food intake was perturbed in both cases, necessitating supplementation via gastrostomy for individual 2 from the age of 3 years onward.

At the time of their last clinical evaluations at the ages of 16 and 10 years, respectively, individuals 1 and 2 had microcephaly, protruding eyes, and mandibular hypoplasia along with hypodontia (Figures 1A–1C). Prominent veins and a paucity of subcutaneous fat were notable in both, although these characteristics were more pronounced in individual 1. Skeletal features included wormian bones and hypoplastic clavicles, as well as reduced bone mineral density in the case of individual 1. Both had normal cognition but cerebellar intention tremor and mildly reduced muscle strength. Cranial MRIs of individual 1 had revealed white-matter hyperintensities of the peritrigonal area at the age of 5 and diffuse hyperintensities of the cerebral white matter at the age of 14, indicating progressive hypomyelination (Figure S1A). A cranial MRI of individual 2 at the age of 22 months did not show white-matter irregularities. Whereas, in the case of individual 2, growth hormone (GH) therapy initiated at age 4 substantially improved linear growth, GH therapy beginning at age 11 seemed to be ineffective for individual #1 and was discontinued after 19 months because of newly occurring insulin resistance. Repeated ophthalmological examination of both individuals revealed no significant pathologies.

Family history of both individuals was negative for similar symptomatology, and there was no history of consanguinity. More detailed descriptions of the clinical histories of individuals 1 and 2 can be found in the Supplemental Data. A compilation of the phenotypes in

Table 1. Compilation of Phenotypic Features of Both Individuals in Comparison to Those of Classical HGPS, NGPS, and PYCR1-Related Autosomal-Recessive Cutis Laxa

Disorder		LEMD2-Associated Nuclear EnveloThy, Individual 1	LEMD2-Associated Nuclear EnveloThy, Individual 2	Hutchinson-Gilford Progeria Syndrome	Néstor-Guillermo Progeria Syndrome	PYCR1-Related Autosomal Recessive Cutis Laxa
Gene, Protein(s)		LEMD2, LEM domain-containing protein 2	LEMD2, LEM domain-containing protein 2	LMNA, lamin A and C	BANF1, barrier-to-autointegration factor	PYCR1, pyrroline-5-carboxylate reductase 1
Facial Features	Triangular face	+++	+++	++	+	+++
	Prominent eyes	+++	+++	+++	+++	++
	Crooked nose, septum deviation	+++	+	++	+++	-
	Mandibular hypoplasia	+++	++	++	+++	-
Skin, Hair, Nails, and	Juvenile alopezia	-	-	+++	++	-
Subcutaneous Tissue	Loss of eyebrows	-	-	+++	-	-
	Thin skin, prominent veins	+	++	+++	+++	++
	Cutis laxa	-	-	-	-	+++
	Patchy hyperpigmentation	-	-	+	+	-
	Dystrophic nails	-	-	+	+	-
	Generalized lipoatrophy	++	+	+++	++	-
Skull and Dentures	Microcephaly	+++	++	++	-	+
	Wormian bones	+++	+++	-	-	+
	Open cranial sutures	-	-	-	+++	-
	Dental crowding	+++	+++	-	+++	-
	Supernumerary teeth	+++	+++	-	-	-
	Delayed dentition	+++	+++	+	-	-
Skeletal Features	Short stature	+++	-[a]	+++	+++	+
	Low bone density	++	-	++	+++	++
	Hypoplastic clavicles	++	++	++	+++	-
	Osteolytic foci	-	-	++	+++	-
	Joint hyperlaxity	-	-	-	-	++
	Limited joint mobility	-	-	++	+++	-
Other Features	IUGR	+	-	-	-	++
	Atherosclerosis	-	-	+++	-	-
	Intention tremor	++	++	-	-	-
	Developmental delay	-	-	-	-	++

Compilation of phenotypic features of both individuals in comparison to those of classical Hutchinson-Gilford Progeria Syndrome (HGPS), the phenotypically over-lapping Néstor-Guillermo Progeria Syndrome (NGPS), and PYCR1-related autosomal-recessive cutis laxa, to which a high facial similarity was demonstrated upon image analysis. The novel LEMD2-associated phenotype has a characteristic and distinctive pattern of features involving several organ systems but still displaying the highest similarity to the nuclear envelopathies NGPS and HGPS.
[a]Normal height (32th percentile) after GH-Treatment beginning at the age of 4 years.

comparison to those associated with HGPS,[2] the segmental progeroid disorders known as Néstor-Guillermo progeria syndrome (MIM: 614008),[4] and PYCR1-related auto-somal-recessive cutis laxa (MIM: 612940),[10] to which high facial similarity to individual 1 and 2 was calculated by FaceNet (see Figure S2B), is shown in Table 1. Of note,

PYCR1-related autosomal-recessive cutis laxa has recently been reported to display significant clinical similarities to HGPS, predominantly in the first years of life.[11]

Individual 1 had undergone molecular genetic testing of LMNA and ZMPSTE24 (MIM: 606480) prior to recruitment for the research study, and individual 2 had received
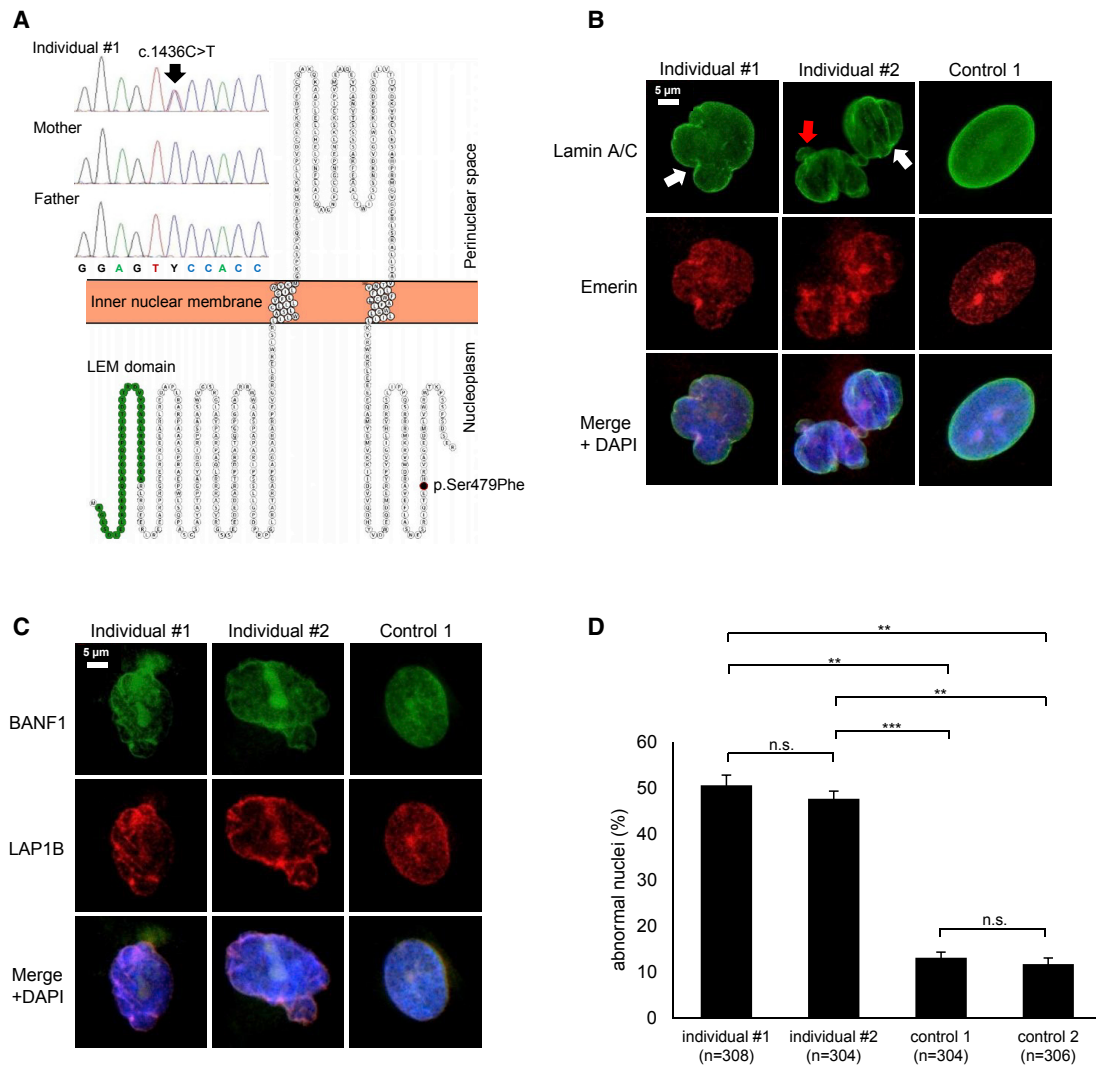
Figure 2. Cellular Phenotype of Patient Fibroblasts

(A) Sanger Sequencing of exon 9 of *LEMD2* in individual 1 and his parents shows the c.1436C>T *de novo* mutation in a heterozygous state. The resulting amino acid change p.Ser479Phe is indicated within the amino acid sequence of LEMD2. The N-terminal LEM domain (IPR011015), which mediates DNA binding through barrier-to-autointegration factor (BAF) proteins, and the mutated amino acid position within the C-terminal nucleoplasmic domain are highlighted. The image was generated with the Protter website (see Web Resources).

(B and C) Nuclear abnormalities in patient fibroblasts compared to control fibroblasts can be seen across multiple different IF stainings of NE proteins: stainings of lamin A and C, Emerin, BANF1, and LAP1B (TOR1AIP1) reveal irregular shapes, "blebbing" (red arrow), and invaginations (white arrows) of the nuclear membrane in patient fibroblasts (representative images).

(D) Quantification of nuclear anomalies in primary fibroblasts shows a significantly higher percentage of abnormal nuclei in both individuals' fibroblasts as compared to cells from healthy controls. The combined results (mean $\pm$ SEM) of three separate experiments are shown (** = $p \leq 0.01$, *** = $p \leq 0.001$).

testing for *LMNA* and an SNP chromosomal microarray. Because these tests could not provide an explanation for the unusual phenotypes, trio-exome sequencing of both individuals and their parents was performed (see Supplemental Materials and Methods for details). The German and Norwegian research groups independently identified the *de novo* missense variant c.1436C>T, (p. Ser479Phe) (GenBank: NM_181336.3) in *LEMD2*, which results in the substitution of serine with phenylalanine at position 479 (Figure 2A), as potentially causative. When both groups came into contact through GeneMatcher, a web-

based tool for researchers and clinicians working on the same genes,[12] the fact that both individuals carried the identical *de novo* variant as well as striking similarities in appearance and medical history became apparent. Given the rarity of *de novo* mutations in the protein-coding part of the genome, the causal role of the *LEMD2* variant was established at this stage; the likelihood of finding the same *de novo* variant—by chance—in two individuals with a distinct and similar phenotype is negligibly low.[13] We estimate this likelihood to be approximately 1 in 60 million (see Supplemental Note 2 for further details).

As is to be expected for the causal variant in an ultra-rare disease, the variant c.1436C>T was absent from the ExAC and gnomAD databases. The algorithms MutationTaster, PolyPhen-2, and SIFT predicted the variant to be disease causing, probably damaging, and damaging, respectively. The CADD score was calculated as 32.000, a value strongly suggesting deleteriousness of a variant. The fact that neither affected individuals carried a second rare variant in LEMD2 suggests an autosomal-dominant mode of inheritance of the phenotype. Of note, the Hutterite founder mutation p.Leu13Arg (c.38T>G), which causes an autosomal-recessive type of juvenile-onset cataract (accompanied by sudden cardiac death at an early age in some individuals), affects an amino acid within the N-terminal LEM domain, whereas the variant p.Ser479Phe (c.1436C>T) is located at the C terminus of LEMD2 (Figure 2A). Viewed from a more global perspective, the number of observed LEMD2 missense variants (irrespective of their allele frequency) in the ExAC and gnomAD databases is lower than the number of theoretically expected variants[14] for this gene: the missense Z score taken from the constraint metric for LEMD2 is 3.65 in ExAC (203.3 expected and 97 observed missense variants) and 1.98 in gnomAD (226.2 expected and 144 observed), indicating a relatively high degree of evolutionary intolerance to missense variation.

Considering the well-established role of LEMD2 in the maintenance of nuclear membrane morphology[15] and LEMD2's previously reported association with lamin A and emerin,[6,16] we performed immunofluorescence analysis of patient-derived fibroblasts and fibroblasts of two healthy control individuals (see Supplemental Material and Methods for experimental details). This analysis, through the use of four well-established integral members of the NE (lamin A and C, emerin, BANF1, and LAP1B), revealed significantly increased numbers of abnormal nuclei in both patient cell lines when compared to the control cell lines (Figures 2B–2D), although all four investigated proteins showed correct intracellular localization. These results were similar to previous findings in HGPS cells.[17] Notably, although we observed some nuclear blebs and complex nuclear abnormalities, the gross majority (~80%–90%) of nuclear abnormalities in fibroblasts from both individuals were invaginations of the NE (Figure 2B, white arrows). These data recapitulate findings in LEMD2-deficient cell lines,[15] providing additional evidence for the pathogenic nature of the here identified LEMD2 variant.

We further tried to evaluate the localization of LEMD2 in patient fibroblasts compared to control fibroblasts. However, both tested antibodies gave unspecific staining (data not shown). We therefore turned to a U2OS transient transfection system. Overexpression of wild-type (WT) and mutant LEMD2-FLAG constructs in U2OS cells led to frequent formation of "patches" of LEMD2 within the nuclear periphery, as previously described.[6] However, transfection with small amounts of plasmid DNA (125 ng/mL)

resulted in smooth distribution and good overall co-localization with lamin A and C in the case of WT LEMD2, whereas patchy accumulations remained predominant in cells transfected with mutant LEMD2 (Figures 3A and 3B). This visual impression was validated by quantitative co-localization analysis (Figure 3C). Both WT and mutant LEMD2-FLAG constructs were largely restricted to the nucleus, and immunoblotting showed no differences in spatial distribution within the cell (Figure 3D). The notion that mutant LEMD2 accumulates in direct contact with the nuclear lamina was further supported by co-localization of mutant and WT LEMD2-FLAG constructs with the lamina-associated proteins emerin and BANF1 (Figures 3E–3G). The preferential formation of "patches" suggests a mislocalization of mutant LEMD2 within the NE, which is in accordance with the observation of invaginations of the nuclear membrane in patient fibroblasts.[18] Taken together, these data suggest that this disorder is a nuclear envelopathy on the cellular level.

When the German group analyzed frontal photographs of individual 1 in early 2017 by using Face2Gene, the well-established nuclear envelopathy HGPS was listed among the most likely diagnoses, and later image analysis of individual 2 produced similar results. We considered it an intriguing finding that DeepGestalt,[19] this website's deep convolutional neural network, which is commonly used as an aid to diagnosis in patients with known disorders, correctly assigned a novel syndrome to a group of syndromes for which facial similarity suggested an overlapping pathogenesis. Both individuals' DeepGestalt similarity scores for known disorders, as well as the overlap between the individual DeepGestalt similarity scores with regard to each disorder, are shown on a radar plot in Figure S2A. DeepGestalt was trained via a supervised learning process to assign patient pictures to a set of 216 syndromes it had "learned" to recognize beforehand. In order to visualize the results of the DeepGestalt analysis, we generated composite masks of both syndromes. These masks demonstrate facial similarities and differences between individuals 1 and 2 and patients with HGPS (Figure 1D).

Because the genetic disorder of individual 1 and 2 was previously unknown, we decided to also investigate the similarity between both cases in an unsupervised, data-driven way. To do this, we clustered patient photos by using the neural network FaceNet,[20] which was initially trained to recognize intra-personal similarity in a large dataset of unconstrained facial photos (i.e., in a non-medical context). After being trained to recognize intra-syndromal similarity in a group of 265 individuals with facial abnormalities from 66 different monogenic syndromes that had been selected from the PEDIA cohort (T.-C. H. et al., unpublished data), FaceNet found individuals 1 and 2 to be not only almost the most similar among 34,980 random pairwise comparisons but also more similar than most individuals inside other disease entities, including some related individuals (Figure 1E).
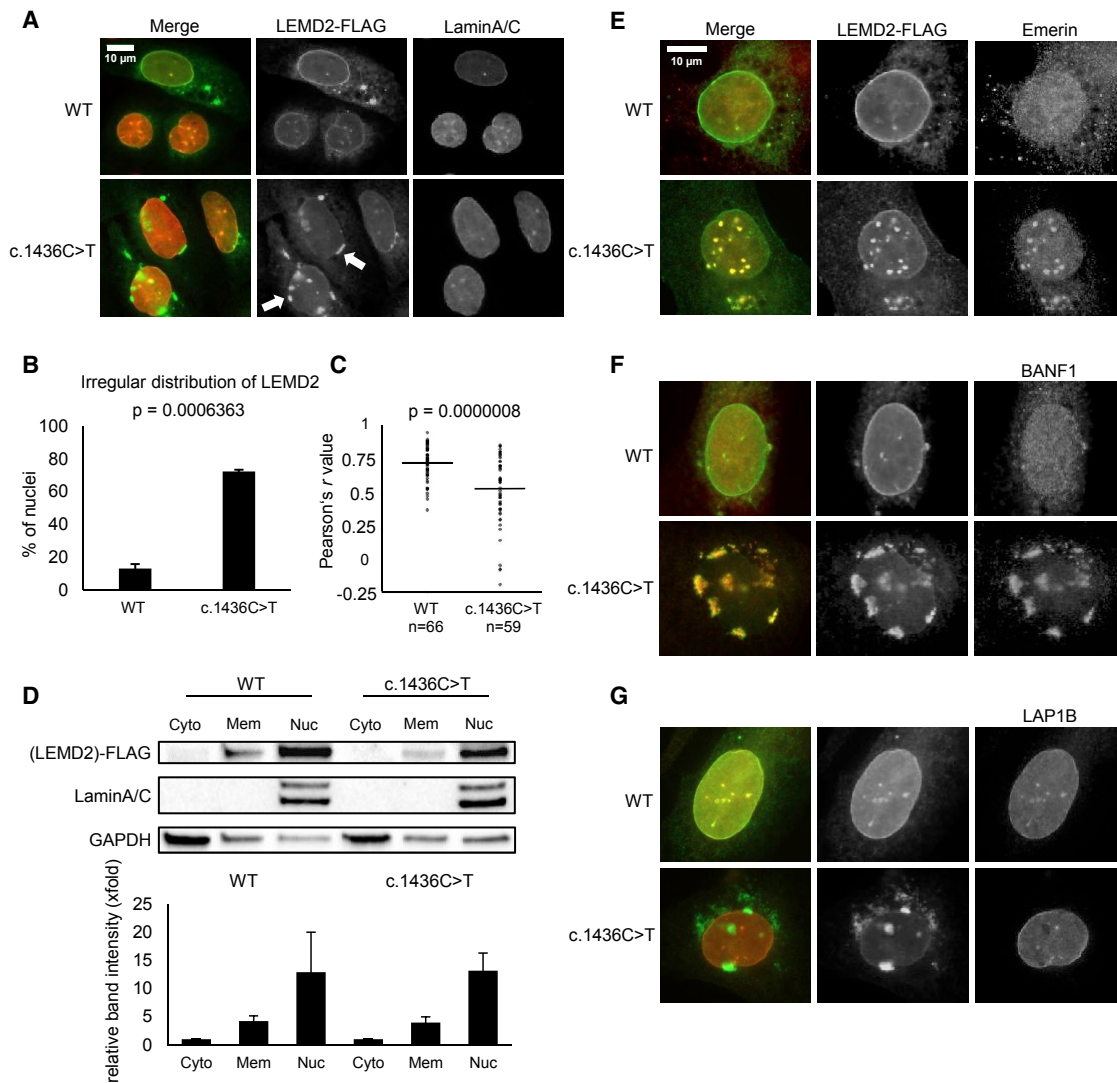
Figure 3. Immunofluorescence Analysis of Mutant *LEMD2* in U2OS Cells

(A) Irregular nuclear localization of mutant LEMD2 during transient overexpression with pCMV6-*LEMD2*-FLAG constructs containing either the *LEMD2* WT or mutant (c.1436C>T) sequence in U2OS cells. 48 h after transfection, immunofluorescence staining of FLAG (green) and lamin A and C (red) reveals patchy accumulations of mutant LEMD2 (white arrows), whereas the WT construct shows more even distribution within the nuclear lamina.

(B) Manual comparison of cells transfected with WT and mutant *LEMD2* constructs shows a higher percentage of abnormal distribution in cells transfected with mutant *LEMD2*. Nuclear distribution of LEMD2 was characterized as normal or irregular, and the percentage of nuclei with irregular distribution was calculated. Combined results (mean ± SEM) of three separate experiments are shown. (A total of 225 cells transfected with WT and 209 cells transfected with mutant *LEMD2* were analyzed.)

(C) Quantitative colocalization analysis of WT and mutant LEMD2 demonstrates disturbed colocalization of mutant LEMD2 and lamin A and C. Nuclei of cells transfected with both constructs (pCMV6-*LEMD2*-FLAG WT and mutant) were delimited and analyzed with the ImageJ software (Coloc2 plugin). The Pearson correlation coefficients of LEMD2-FLAG and lamin A and C of each individual nucleus, as well as averages and p values (according to the Wilcoxon rank sum test with continuity correction), are shown.

(D) WT and mutant LEMD2 are both predominantly localized within the nucleus. The Subcellular Protein Fractionation Kit (Thermo Fisher) was used to collect fractionated cell lysates 48 h after transfection with WT and mutant *LEMD2* constructs. Immunoblot analysis revealed similar distribution patterns of both WT and mutant LEMD2 within cytoplasm (Cyto), membranes (Mem), and nuclear (Nuc) fractions. Comparison of band intensity normalized for total protein (shown as x-fold change with regard to the normalized band intensity of the cytoplasm fraction) from three separate experiments (mean ± SEM) did not reveal significant differences between WT and mutant constructs.

(E–G) Proteins of the nuclear lamina display different localization patterns relative to LEMD2-FLAG constructs. U2OS cells were transfected through the use of pCMV6-*LEMD2*-FLAG constructs (WT or mutant), and immunofluorescence stainings were performed 48 h after transfection. Whereas emerin and BANF1 colocalize with irregular nuclear accumulations of mutant LEMD2 (E and F), LAP1B does not (G).

A close-up of the dendrogram of the cophenetic distances of the clustering analysis is shown in Figure S2B. We concluded that FaceNet could determine, solely on the basis of the facial phenotype, that individuals 1 and 2 were probably affected by the same syndrome and that DeepGestalt recognized that this particular phenotype was similar to that of the family of early-onset segmental progeroid syndromes. More details about the techniques, including the cluster analysis, can be found in the Supplemental Data.

The LEMD2-associated disorder presented in this report displays a distinct combination of symptoms and has phenotypic overlap with other established segmental progeroid syndromes, such as HGPS, mandibuloacral dysplasia (MIM: 248370, 608612), and Néstor-Guillermo Progeria syndrome. However, it is important to note the better overall prognosis of the LEMD2-associated disorder when compared to HGPS: individual 1 is now 19 years of age and without clinical signs of atherosclerosis, and neither patient currently presents with life-limiting symptoms.

The available data on LEMD2,[6–8,15,16,21] combined with the individuals' phenotypes and our own observations in vitro, strongly support our hypothesis of an LEMD2-associated syndromic nuclear envelopathy. The maintenance of nuclear morphology is a well-documented function of LEMD2 in vitro: it is required for reformation and closure of the NE during anaphase in mitotic yeast and HeLa cells,[21] high-level overexpression of LEMD2 led to the formation of finger-type intrusions in the nuclear membrane of HeLa cells,[6] and invaginations and lobulations of the NE have been reported upon siRNA-mediated depletion of LEMD2 in HeLa and U2OS cells.[15] Thus, the nuclear anomalies in patient fibroblasts (Figure 2B) recapitulate the findings in the above-mentioned LEMD2-deficient cell lines,[15] and these anomalies might be attributable to the observed irregular accumulations of mutant LEMD2 in the NE. Besides having an established role in the maintenance of NE integrity, LEMD2 has additionally been implicated in several other mutually non-exclusive biological processes and signaling pathways, including telomere anchoring and heterochromatin silencing,[22] DNA replication,[23] and MAPK cascade and ERK signaling.[24] However, elucidation of the impact of the p.Ser479Phe variant in these processes as well as those linked to other progeroid syndromes will require further work and will be the main focus of our future studies.

Traditionally, recognition and classification of facial dysmorphology are among the key skills of a clinical geneticist. What used to require decades of training and clinical experience is now being supplemented by the introduction of image-analysis algorithms into clinical practice. For diagnoses with sufficiently distinct facial phenotypes, algorithms employing machine learning generally fare well in prioritizing potential diagnoses, sometimes with even better accuracy than that of seasoned geneticists.[19,25,26] Another not-yet-exploited utility of these tech-nologies is their potential to detect similarities of novel phenotypes to known disorders and to thereby identify a potential molecular etiology, such as the disruption of a signaling pathway or a specific cellular compartment. In our case, DeepGestalt recognized the facial similarity of individuals with the LEMD2 mutation c.1436C>T to individuals with another nuclear envelopathy, HGPS. Furthermore, FaceNet, when trained accordingly to detect intra-syndromal similarity, correctly perceived that these two individuals have highly similar facial dysmorphology. Both neural networks thus provided supporting evidence for gene identification in our particular case. We suggest that, in particular, FaceNet's ability to form groups within a cohort of syndromic patients constitutes a hypothesis-free approach to syndrome identification and patient matching and that this approach could even help to identify individuals with ultra-rare or previously unknown disorders (e.g. via a web-based patient-matching tool). Such image-analysis algorithms could be employed in the detection of syndromic "kinship" between non-related individuals and in the delineation of syndrome families to facilitate the discovery of new genetic disorders and underlying mutations, a concept that was suggested almost 15 years ago.[27]

In summary, we show that a recurrent de novo mutation in LEMD2 causes a syndromic nuclear envelopathy with a relatively good prognosis, and we suggest that the use of artificial intelligence could facilitate the discovery of new genetic disorders.

Web Resources

CADD: Combined Annotation Dependent Depletion, https://cadd.gs.washington.edu/

ExAC, http://exac.broadinstitute.org/

Face2Gene, https://www.face2gene.com/

GeneReviews, Bonne, G., Leturcq, F., Ben Yaou, R. (2004). Emery-Dreifuss Muscular Dystrophy, https://www.ncbi.nlm.nih.gov/books/NBK1436/

GeneReviews, Gordon, L.B., Brown, W.T., Collins, F.S. (2003). Hutchinson-Gilford Progeria Syndrome, https://www.ncbi.nlm.nih.gov/books/NBK1121/

gnomAD, https://gnomad.broadinstitute.org/

Leiden Open (Source) Variation Database (LOVD), http://www.lovd.nl/3.0/home

MutationTaster, http://www.mutationtaster.org/

Online Mendelian Inheritance in Man (OMIM), https://www.omim.org/

PolyPhen-2, http://genetics.bwh.harvard.edu/pph2/

Protter, http://wlab.ethz.ch/protter

SIFT, https://sift.bii.a-star.edu.sg/

References

1. Somech, R., Shaklai, S., Amariglio, N., Rechavi, G., and Simon, A.J. (2005). Nuclear envelopathies–raising the nuclear veil. Pediatr. Res. 57, 8R–15R.

2. Merideth, M.A., Gordon, L.B., Clauss, S., Sachdev, V., Smith, A.C., Perry, M.B., Brewer, C.C., Zalewski, C., Kim, H.J., Solomon, B., et al. (2008). Phenotype and course of Hutchinson-Gilford progeria syndrome. N. Engl. J. Med. 358, 592–604.

3. Davidson, P.M., and Lammerding, J. (2014). Broken nuclei–lamins, nuclear mechanics, and disease. Trends Cell Biol. 24, 247–256.

4. Cabanillas, R., Cadiñanos, J., Villameytide, J.A., Pérez, M., Longo, J., Richard, J.M., Alvarez, R., Durán, N.S., Illán, R., González, D.J., and López-Otín, C. (2011). Néstor-Guillermo progeria syndrome: a novel premature aging condition with early onset and chronic development caused by BANF1 mutations. Am. J. Med. Genet. A. 155A, 2617–2625.

5. Hellemans, J., Preobrazhenska, O., Willaert, A., Debeer, P., Verdonk, P.C., Costa, T., Janssens, K., Menten, B., Van Roy, N., Vermeulen, S.J., et al. (2004). Loss-of-function mutations in LEMD3 result in osteopoikilosis, Buschke-Ollendorff syndrome and melorheostosis. Nat. Genet. 36, 1213–1218.

6. Brachner, A., Reipert, S., Foisner, R., and Gotzmann, J. (2005). LEM2 is a novel MAN1-related inner nuclear membrane protein associated with A-type lamins. J. Cell Sci. 118, 5797–5810.

7. Barrales, R.R., and Braun, S. (2016). Chromatin binding and silencing: Two roles of the same protein Lem2. Microb. Cell 3, 185–188.

8. Thanisch, K., Song, C., Engelkamp, D., Koch, J., Wang, A., Hallberg, E., Foisner, R., Leonhardt, H., Stewart, C.L., Joffe, B., and Solovei, I. (2017). Nuclear envelope localization of LEMD2 is developmentally dynamic and lamin A/C dependent yet insufficient for heterochromatin tethering. Differentiation 94, 58–70.

9. Boone, P.M., Yuan, B., Gu, S., Ma, Z., Gambin, T., Gonzaga-Jauregui, C., Jain, M., Murdock, T.J., White, J.J., Jhangiani, S.N., et al. (2015). Hutterite-type cataract maps to chromosome 6p21.32-p21.31, cosegregates with a homozygous mutation in LEMD2, and is associated with sudden cardiac death. Mol. Genet. Genomic Med. 4, 77–94.

10. Dimopoulou, A., Fischer, B., Gardeitchik, T., Schröter, P., Kayserili, H., Schlack, C., Li, Y., Brum, J.M., Barisic, I., Castori, M., et al. (2013). Genotype-phenotype spectrum of PYCR1-related autosomal recessive cutis laxa. Mol. Genet. Metab. 110, 352–361.

11. Lessel, D., Ozel, A.B., Campbell, S.E., Saadi, A., Arlt, M.F., McSweeney, K.M., Plaiasu, V., Szakszon, K., Szőllős, A., Rusu, C., et al. (2018). Analyses of LMNA-negative juvenile progeroid cases confirms biallelic POLR3A mutations in Wiedemann-Rautenstrauch-like syndrome and expands the phenotypic spectrum of PYCR1 mutations. Hum. Genet. 137, 921–939.

12. Sobreira, N., Schiettecatte, F., Boehm, C., Valle, D., and Hamosh, A. (2015). New tools for Mendelian disease gene identification: PhenoDB variant analysis module; and GeneMatcher, a web-based tool for linking investigators with an interest in the same gene. Hum. Mutat. 36, 425–431.

13. Conrad, D.F., Keebler, J.E., DePristo, M.A., Lindsay, S.J., Zhang, Y., Casals, F., Idaghdour, Y., Hartl, C.L., Torroja, C., Garimella, K.V., et al.; 1000 Genomes Project (2011). Variation in genome-wide mutation rates within and between human families. Nat. Genet. 43, 712–714.

14. Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al.; Exome Aggregation Consortium (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature 536, 285–291.

15. Ulbert, S., Antonin, W., Platani, M., and Mattaj, I.W. (2006). The inner nuclear membrane protein Lem2 is critical for normal nuclear envelope morphology. FEBS Lett. 580, 6435–6441.

16. Huber, M.D., Guan, T., and Gerace, L. (2009). Overlapping functions of nuclear envelope proteins NET25 (Lem2) and emerin in regulation of extracellular signal-regulated kinase signaling in myoblast differentiation. Mol. Cell. Biol. 29, 5718–5728.

17. Eriksson, M., Brown, W.T., Gordon, L.B., Glynn, M.W., Singer, J., Scott, L., Erdos, M.R., Robbins, C.M., Moses, T.Y., Berglund, P., et al. (2003). Recurrent de novo point mutations in lamin A cause Hutchinson-Gilford progeria syndrome. Nature 423, 293–298.

18. Vigouroux, C., Auclair, M., Dubosclard, E., Pouchelet, M., Capeau, J., Courvalin, J.C., and Buendia, B. (2001). Nuclear envelope disorganization in fibroblasts from lipodystrophic patients with heterozygous R482Q/W mutations in the lamin A/C gene. J. Cell Sci. 114, 4459–4468.

19. Gurovich, Y., Hanani, Y., Bar, O., Nadav, G., Fleischer, N., Gelbman, D., Basel-Salmon, L., Krawitz, P.M., Kamphausen, S.B., Zenker, M., et al. (2019). Identifying facial phenotypes of genetic disorders using deep learning. Nat. Med. 25, 60–64.

20. Schroff, F., Kalenichenko, D., and Philbin, J. (2015). FaceNet: A Unified Embedding for Face Recognition and Clustering. arXiv, arXiv:1503.03832. https://arxiv.org/abs/1503.03832.

21. Gu, M., LaJoie, D., Chen, O.S., von Appen, A., Ladinsky, M.S., Redd, M.J., Nikolova, L., Bjorkman, P.J., Sundquist, W.I., Ullman, K.S., and Frost, A. (2017). LEM2 recruits CHMP7 for ESCRT-mediated nuclear envelope closure in fission yeast and human cells. Proc. Natl. Acad. Sci. USA *114*, E2166–E2175.

22. Barrales, R.R., Forn, M., Georgescu, P.R., Sarkadi, Z., and Braun, S. (2016). Control of heterochromatin localization and silencing by the nuclear membrane protein Lem2. Genes Dev. *30*, 133–148.

23. Xu, Y.J. (2016). Inner nuclear membrane protein Lem2 facilitates Rad3-mediated checkpoint signaling under replication stress induced by nucleotide depletion in fission yeast. Cell. Signal. *28*, 235–245.

24. Tapia, O., Fong, L.G., Huber, M.D., Young, S.G., and Gerace, L. (2015). Nuclear envelope protein Lem2 is required for mouse development and regulates MAP and AKT kinases. PLoS ONE *10*, e0116196.

25. Dudding-Byth, T., Baxter, A., Holliday, E.G., Hackett, A., O'Donnell, S., White, S.M., Attia, J., Brunner, H., de Vries, B., Koolen, D., et al. (2017). Computer face-matching technology using two-dimensional photographs accurately matches the facial gestalt of unrelated individuals with the same syndromic form of intellectual disability. BMC Biotechnol. *17*, 90.

26. Ferry, Q., Steinberg, J., Webber, C., FitzPatrick, D.R., Ponting, C.P., Zisserman, A., and Nellåker, C. (2014). Diagnostically relevant facial gestalt information from ordinary photos. eLife *3*, e02020.

27. Brunner, H.G., and van Driel, M.A. (2004). From syndrome families to functional genomics. Nat. Rev. Genet. *5*, 545–551.

Supplemental Data

The Discovery of a *LEMD2*-Associated Nuclear

Envelopathy with Early Progeroid Appearance Suggests

Advanced Applications for AI-Driven Facial Phenotyping

Felix Marbach, Cecilie F. Rustad, Angelika Riess, Dejan Đukić, Tzung-Chien Hsieh, Itamar Jobani, Trine Prescott, Andrea Bevot, Florian Erger, Gunnar Houge, Maria Redfors, Janine Altmueller, Tomasz Stokowy, Christian Gilissen, Christian Kubisch, Emanuela Scarano, Laura Mazzanti, Torunn Fiskerstrand, Peter M. Krawitz, Davor Lessel, and Christian Netzer

Supplementary appendix

Table of contents

Supplemental Note #1: Case Reports

Medical history and clinical findings of Individual #1 and #2

Individual #1

Individual #1 was born at 36 weeks gestation, weighing 1910g (P8, Z-1.44), with a length of 44cm (P35, Z-0.4) and head circumference of 31cm (P16, Z-1.01) (Percentiles at birth according to *CDC growth charts*; www.cdc.gov/growthcharts)[1]. Intrauterine growth restriction had been noticed on prenatal ultrasound. Unspecified respiratory complications necessitated mechanical ventilation during the first days of life. He had normal psychomotor development during early childhood, walking freely at the age of 18 months. Dentition was severely delayed, with the first tooth erupting at the age of three. At the age of five years, he received a cranial MRI demonstrating hyperintensities of the peritrigonal white matter. Neurologic examination at the time was normal with the exception of a fine tremor of the lower extremities and hands. A second MRI with magnetic resonance spectroscopy (MRS) was performed at age 11, which revealed diffuse hypomyelination of the white matter. Individual #1 attended school in Italy and Germany, successfully graduating from secondary school. At 11 years of age, growth hormone (GH) therapy was administered for 19 months albeit normal GH levels without satisfactory results and had to be discontinued upon newly occurring insulin resistance: Elevated HOMA-IR (Homeostatic Model Assessment for Insulin Resistance) values were noted after 6 months of therapy, which only receded to the normal range two years after cessation of GH therapy.

Lipid profiles (cholesterol, HDL, triglycerides) were repeatedly determined as normal between 11 and 15 years of age. LDL was determined at age 15, with normal results.

Echocardiography and 24h-Holter monitoring at the age of 14 were both normal. The most recent echocardiography performed at the age of 17 years demonstrated normal atrial and ventricular dimensions, septal hypertrophy, a functionally bicuspid aortic valve, and a low-grade tricuspid and pulmonary valve insufficiency. Recent electrocardiograms showed no pathologies apart from an incomplete right bundle branch block and signs of septal hypertrophy.

Individual #1 suffered fractures of the femur and humerus at the age of 13 and 15, respectively, due to inadequate traumata. Orthopedic treatment of the femur by internal fixation was complicated by perforation of Prévot nails and breaking of a plate and screw fixation, necessitating osteotomy. Reduced bone mineral density was noticed by quantitative computed tomography. Low appetite and preference of small portion sizes, resulting in malnutrition, was treated with hypercaloric supplements.

Neurologic examination at the age of 14 revealed muscular weakness and cerebellar (intention-) tremor, which intensified in stressful situations. Intelligence testing showed normal cognition. A third MRI of the neurocranium demonstrated diffuse white matter hyperintensities in T2, indicating widespread hypomyelination (Fig, S1A). Delayed gastric passage and a "fish-hook stomach" conformation was noted upon barium esophagography. Skeletal features included severe growth retardation with a height of 139 cm (<P1, Z-4.95), hypoplastic clavicles and microcephaly with a head circumference of 47.2 cm (<P1, Z-6.15) at the age of 16 (Percentiles in adolescence according to *Kromeyer-Hauschild et al., 2001*)[2]. Dysmorphic features of the skull included wormian bones, small palate, nasal septum deviation, absence of the right and hypotrophy of the left maxillary sinus. He had shallow orbits, a beaked nose, mandibular hypoplasia and severe hypodontia (Fig. 1B, S1B). The individual's denture consisted predominantly of primary teeth, the only permanent teeth being two incisors which erupted at the age of 12. The upper and lower jaws are crowded with unerupted teeth (Fig. 1C). Ophthalmological examination at age 16 did not reveal cataracts or retinal abnormalities.

Individual #2

Individual #2 was conceived by In-vitro-fertilization. Pregnancy was normal until premature rupture of membranes at 28 weeks gestation. He was born at 32 weeks with a weight of 1675g (P33, Z-0.44), length of 40.5 cm (P54, Z-0.09) and a head circumference of 26cm (P5, Z-1.66). Pulmonary hypertension was treated with high-frequency oscillation ventilation for the first three days. Neonatal investigations included a normal cerebral ultrasound initially, followed by an MRI scan showing a left hemispheric subcortical hemorrhage with edema and two normal echocardiograms. A second cerebral MRI at 22 months of age revealed a thin rim of anterior and a normal posterior pituitary. Both MRI scans did not indicate hypomyelination. The boy breastfed well for the first four months of life. Gradual transition to solid foods from age four months was uneventful. Beginning at two years, he became disinterested in food. Growth retardation persisted in the absence of malabsorption or endocrine anomalies. Percutaneous gastrostomy placed at age three improved growth to some degree, but a persistent lack of catch-up growth initiated the decision for GH treatment at the age of four, which continued to date of this report. At the age of seven, he was still dependent on gastrostomy for nutrition. Individual#2 had a thorough cardiac examination at the age of seven, which did not reveal any pathologies. As he showed no eruption of teeth, surgical excision of regular and supernumerary teeth, as well as exposure of permanent teeth were performed between the age of seven and nine years (Fig 1C, S1C). At ten years he ate solid food, but caloric intake had to be supplemented via gastrostomy. Cognitive, motor and social development were normal. He attended regular school, taking part in all activities including physical education, although he had less muscle strength than his peers.

From an early age a symmetric triangular face and large forehead with visible superficial veins were noted. Additional findings in childhood included a small mouth, smooth philtrum, shallow orbits and a generalized paucity of subcutaneous fat. Midface hypoplasia and growth retardation, created the illusion of macrocephaly. Some of these features diminished with age, albeit persisting mandibular hypoplasia and shallow orbits at the age of 10 (Fig. 1A). Individual #2 was 138.6cm (P32 Z-0.46) tall and

weighed 28.5 kg (P15 Z-1.04). Neurologic examination was only notable for an intermittent intention tremor.

Supplemental Note #2: Comment on *de novo* SNVs

With an estimated average *de novo* mutation rate of between 1.0×10^{-8} and 2.5×10^{-8} per base pair per generation[3-7], the expected number of protein-coding *de novo* single nucleotide variants (SNV) in a 45Mb human exome is 0.45-1.13 per individual. Assuming the higher *de novo* mutation rate of 2.5×10^{-8}, the likelihood for any single phenotypically similar individual to carry any *de novo* SNV at the identical position by chance alone is therefore at most 2.81×10^{-8}, or about 1 in 35 million. Observing the identical C>T transition has a further reduced probability of 1.62×10^{-8}, or about 1 in 60 million. This provides extremely strong statistical support for the causal role of this *de novo* SNV in these two phenotypically similar individuals.

The number of expected *de novo* SNVs in the human exome were calculated straightforwardly by multiplying the *de novo* mutation rate with the size of the human exome. The likelihood for a second individual to carry the same *de novo* by chance was then closely approximated as

$$p_{same\_dn} = 1 - (1 - r_{dn})^{n_{dn\_exp}}$$

where r_{dn} is the *de novo* mutation rate and n_{dn_exp} the number of expected *de novos* per individual. This assumes no multiple testing, as no larger cohort was systematically sequenced in this study. Since the recruitment of the second patient was facilitated through GeneMatcher, some (non-quantifiable) multiple testing may exist. However, even under the highly unlikely assumption of 100 phenotypically similar progeroid patients in the "GeneMatcher cohort" having undergone trio exome sequencing, the adjusted probability of finding an additional occurrence of the identical de novo among this group would still be minimal (1.62×10^{-6}, or about 1 in 617,000). The likelihood of a *de novo* SNV of a cytosine in a non-CpG context being a C>T transition was estimated at 58% based on previously published data[7].

Figure S1

A — Individual #1, 14 years — T2

B — Individual #1, 16 years

C — Individual #2, 9 years — Individual #2, 10 years
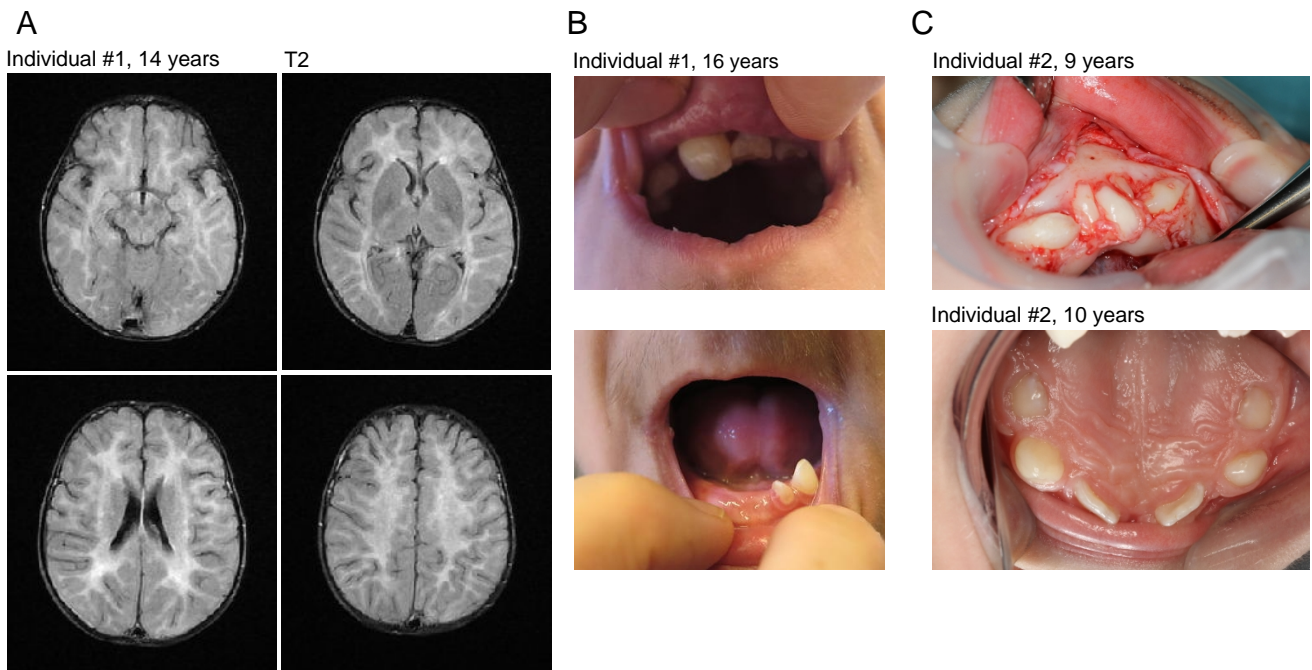
Figure S1

Cranial MRI and dental anomalies.

A: Cranial MRI of individual #1 at the age of 14 years shows widespread diffuse hyperintensities of the cerebral white matter in T2 and T2 FLAIR across all cerebral cross-sections, indicating hypomyelination.

B: Dental status of individual #1 at the age of 16 years. Only a single canine and incisor of the lower, and some incisors of the upper jaw are visible. Note the small oral aperture.

C: Dental features of individual #2. Upper image: Intraoperative image of the lower jaw at the age of 9, when the primary incisors and canines were removed and the permanent tooth germs in the anterior part of the mandible were surgically exposed. Similar surgical treatment was performed in the upper jaw two years earlier. Lower image: Present teeth in the upper jaw at the age of 10 years, three years after surgical exposure of permanent teeth.

Figure S2

A

Gestalt syndrome overlap, patient 1a vs patient 2a

Legend:
- patient 1a
- patient 2a

Radial axis labels (clockwise):
- Wolf-Hirschhorn Syndrome; WHS
- Floating-Harbor Syndrome; FLHS
- Hutchinson-Gilford Progeria Syndrome; HGPS
- Kabuki Syndrome
- Short Syndrome
- Ectrodactyly, Ectodermal Dysplasia, and Cleft Lip/palate Syndrome
- Alagille Syndrome
- Seckel Syndrome
- Holoprosencephaly
- Cutis Laxa, Autosomal Recessive
- Pycnodysostosis
- Silver-Russell Syndrome; SRS
- Williams-Beuren Region Duplication Syndrome
- Rubinstein-Taybi Syndrome
- Loeys-Dietz Syndrome
- Greig Cephalopolysyndactyly Syndrome; GCPS
- Emanuel Syndrome
- Noonan Syndrome
- Crouzon Syndrome
- Prader-Willi Syndrome; PWS
- Geroderma Osteodysplasticum; GO
- Saethre-Chotzen Syndrome; SCS
- Marfan Syndrome; MFS
- Mental Retardation X-Linked 102; MRX102
- Koolen-de Vries Syndrome; KDVS
- Muenke Syndrome; MNKES
- Potocki-Lupski Syndrome; PTLS
- OPITZ-KAVEGGIA SYNDROME; OKS
- Focal Dermal Hypoplasia; FDH
- Hallermann-Streiff Syndrome; HSS

B

Dendrogram leaf labels (top to bottom):
- 226359_LMNA
- 226358_LMNA
- 226381_LMNA
- 226372_LMNA
- 226371_LMNA
- 226354_LMNA
- 226353_LMNA
- 226350_LMNA
- 226399_LMNA
- 79549_PYCR1
- 203981_PYCR1
- 79547_PYCR1
- 79551_PYCR1
- 203979_PYCR1
- patient #2
- patient #1
- 226398_LMNA
- 226363_LMNA
- 226395_LMNA
- 226401_LMNA
- 226356_LMNA
- 226347_LMNA
- 203987_PYCR1
- 203986_PYCR1
- 203988_PYCR1
- 111760_PYCR1
- 111768_PYCR1
- 203983_PYCR1

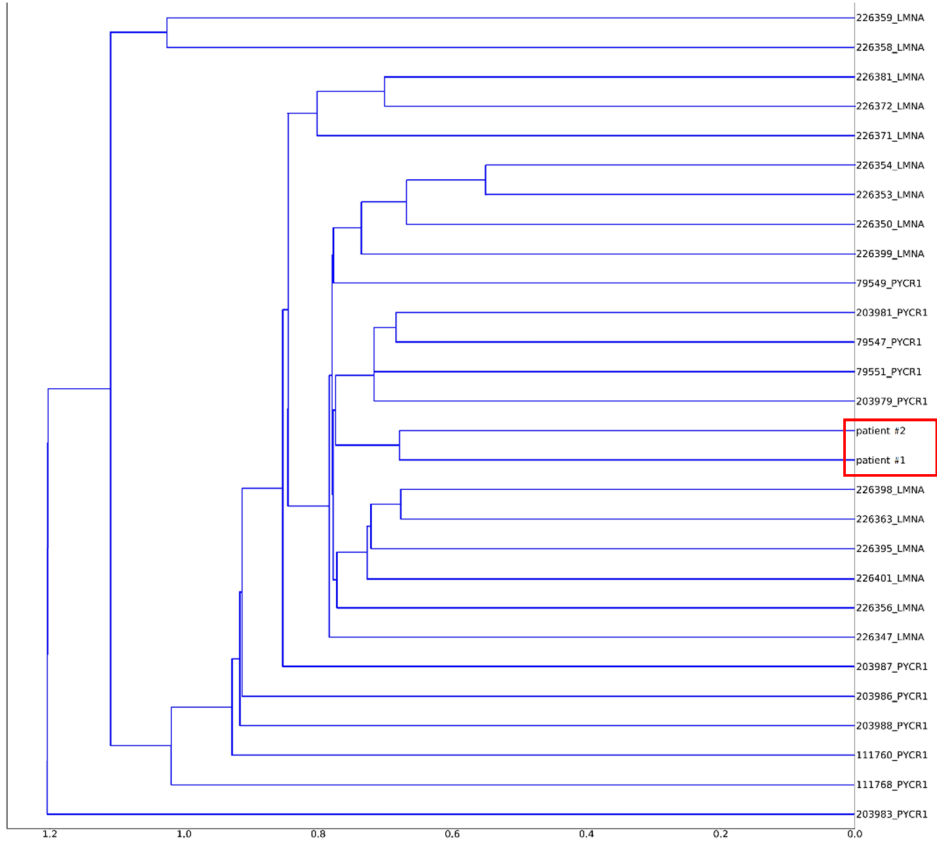X-axis: 1.2 1.0 0.8 0.6 0.4 0.2 0.0

Figure S2

DeepGestalt and FaceNet analysis.

A: Overlap of DeepGestalt similarity scores of individual #1 ("patient 1a", blue) and #2 ("patient 2a", red) to other disorders. The computed facial similarity of each individual to the indicated disorders is visualized as colored areas extending outward on the radicular axes. Purple color indicates overlap.

B: A dendrogram is used to visualize the computed phenotypic "distance" of individual #1 and #2 in a sample containing 265 individuals with 66 different syndromes. The close-up shows both individuals in close proximity to each other, as well as to patients with progeroid disorders (HGPS, *PYCR1*-related autosomal recessive cutis laxa).

Supplemental Material and Methods

Exome sequencing

Cologne:

Library preparation and sequencing in Cologne was performed at the Cologne Center for Genomics (CCG) on an Illumina pipeline as previously described[8].

Bergen:

DNA was extracted from blood using the QIAsymphony system (Qiagen, Hilden, Germany) and exome-sequenced was performed using the SOLID 5500xl platform (Life Technologies, Waltham, MA, USA). Exonic sequences were enriched using the SureSelect Human All Exon v5 kit (Agilent Technologies, Santa Clara, CA, USA) which targets ~21,500 human genes and covers 50Mb of genomic sequence. Color space reads were mapped to the hg19 reference genome using the LifeScope v.2.5 software, yielding on average 120 million mapped reads per exome (75bp paired-end), with about 90% mapped reads on-target. Average depth of coverage within targets was 143. Percent of target bases covered more than 20 times were 89%.

Trio-exome analyses

In order to validate and directly compare the sequencing data of both groups, we proceeded to analyze both trio exome datasets on the same in-house bioinformatics pipeline. NGS sequencing data of individual #2 and his parents were provided by the Department of Clinical Science of the University of Bergen.

FASTQ files were processed on the QIAGEN CLC Biogenomics Workbench using a custom exome analysis workflow. The relevant data processing steps were set as follows: 1. Map Reads to hg19 reference genome, 2. Remove duplicate mapped reads (max. representation of minority sequence: 20%), 3. InDel and Structural Variant Detection (min. variant allele count=5), 4. 2-pass Local Realignment, 5. Variant detection (min. coverage: 16x, min. count: 8, min. variant allele fraction: 20%, ignore broken pairs, minimum base call quality: 26), 6. Remove False Positives (min. allele fraction:

20%, min. fwd/rev-balance: 0.05, min. average base quality: 26). In addition to this, several QC parameters (e.g. coverage information, duplicate read information, etc.) were documented and exported to external storage. A variant call file (.vcf) and binary alignment map were generated with target regions defined as coding regions of all transcribed genomic exons +20bp of adjacent intronic or untranslated sequence.

Variant annotation was performed using the QIAGEN Ingenuity Variant Analysis (IVA) pipeline. The variant filters used to identify potentially pathogenic *de novo* variants in both trio-exome datasets (individual #1+parents, individual #2+parents) consisted of the following filter steps:

1. Filter for variant quality (base call quality & read depth)

2. Filter for rare variants (allele frequency <0.5% in the gnomAD database)[9]

3. Filter for likely deleterious variants based on functional consequence and *in silico* prediction of pathogenicity.

4. Filter for variants only present in the individual's, but not the parents', datasets.

After these filter steps, the number of remaining putative *de novo* variants in both trios was 14 for individual #1, and 15 for individual #2. After manual validation using the Integrative Genomics Viewer (IGV), all variants except the *LEMD2* variant c.1436C>T could be discarded as alignment artefacts in both individuals.

We additionally screened both individuals' exome datasets for heterozygous, hemizygous, compound-heterozygous, or homozygous mutations with clear pathogenicity (ACMG class 5)[10], and variants listed in the HGMD professional database[11]. There were no heterozygous or hemizygous pathogenic variants in autosomal-dominant or X-linked disease-associated genes, nor any homozygous or compound-heterozygous pathogenic variants in autosomal-recessive disease-associated genes in either one of the individual datasets.

Molecular biology

Cell lines

All biological samples from patients were obtained after written informed consent. The study was performed in accordance with the Declaration of Helsinki protocols and approved by the ethics committee of the respective institutions. Primary human dermal fibroblast cultures were established from skin biopsies taken from individuals #1 and #2 and two healthy donors (control1 and control2, aged 34 and 35, respectively). Cells were cultured in Dulbecco's modified Eagle's medium (DMEM), supplemented with 10% fetal bovine serum (Life Technologies, 10270-106), 1% amphotericin B solution (Sigma, A2942) and 1% penicillin-streptomycin solution (Sigma, P4333) under 5% CO_2 and at 37°C. Human bone osteosarcoma epithelial (U2OS) cells were grown on cell culture dishes and coverslips, respectively, utilizing similarly supplemented DMEM. For transient overexpression of *LEMD2*-FLAG constructs, U2OS cells were transfected with 125-500 ng/ml of plasmid-DNA using Lipofectamine® 2000 Transfection Reagent (Thermo Fisher, 11668030) according to the manufacturer's instructions.

Immunofluorescence

Primary fibroblasts and U2OS cells were grown on glass coverslips, fixed with 4% formaldehyde, permeabilized with 0.1% Triton X-100, 0.1% Na-citrate, blocked with 5% BSA in PBS and immunostained with the respective antibodies which were diluted in 5% BSA in PBS. Images were taken using Axioplan 2 imaging microscope (Zeiss, Jena, Germany) and captured using the AxioVision Imaging System (Zeiss, Jena, Germany). Nuclear abnormalities were estimated directly by visual observation at 100x magnification, in at least 100 cells per experiment (primary fibroblasts) and 50 cells/experiment (U2OS cells).

Antibodies

Primary antibodies used in this study were anti-lamin A/C (Santa Cruz, sc-7292), anti-BANF1 (Abcam, ab88464), anti-LAP1B/TOR1AIP1 (Atlas antibodies, HPA050546), anti-emerin (Abcam, ab40688), anti-LEMD2 (Aviva Systems Biology, OAAB14861), anti-LEMD2 (OriGene, TA320148), anti-FLAG (Sigma-Aldrich, F7425) and anti-GAPDH (Sigma-Aldrich, G8795).

Expression constructs

pCMV6-AC-GFP expression vector containing the *LEMD2* coding sequence (NM_181336) was obtained from Origene (RG208535). The missense mutation identified in both individuals was introduced using Quick-Change II site directed mutagenesis kit (Agilent, Waldbronn, Germany), with mutagenic oligonucleotides designed based on the Quick-Change instruction manual. The FLAG-Tag was subsequently added to the C-terminal end of *LEMD2* by PCR-based whole plasmid mutagenesis using Pfu DNA polymerase (Thermo Fisher, EP0501) for plasmid amplification, DpnI (Thermo Fisher, ER1701) for digestion of template DNA, and T4 DNA Ligase (Thermo Fisher, EL0014) for re-ligation of plasmids, using the respective protocols provided by the manufacturer. The following primers were utilized: GCGAGATTACAAGGATGACGACGATAAGTAAACGCGTACGCGGCCGCTCGAG (FLAG forward) and TTACTTATCGTCGTCATCCTTGTAATCTCGCTCTGAGTCAGAGAAGG (FLAG reverse). All constructs were amplified using One Shot™ TOP10 Chemically Competent *E. coli* (Thermo Fisher, C404006) and verified by Sanger sequencing.

Western blotting

U2OS cells were harvested 48h after transfection and processed using the Subcellular Protein Fractionation Kit for Cultured Cells (Thermo Fisher, 78840) according to the manufacturer's instructions. Extracts from cytoplasm, membranes and nucleus were separated by SDS-PAGE (5µg protein/lane) and transferred to PVDF membranes (all components for western blotting were supplied

by Bio-Rad Laboratories Inc., US-CA). PVDF membranes were then stained with primary and secondary antibodies as indicated.

Facial analysis

Creation of composite images

As a first step, faces are located by the Stasm library[12] and 77 landmark points are identified. The positions of the points are averaged out for all of the faces in the set (10 faces included), scaled and aligned in a rectangle in the center of the screen. Delaunay triangles are identified among all the aligned sets of landmark points, creating a mask onto which the faces are then warped and pixels are averaged out, creating a composite mask.

DeepGestalt analysis

The similarity of patients' portrait photos was measured in the clinical face phenotype space (CFPS), a concept first introduced by Ferry et al.[13]. As the clustering efficiency for novel phenotypes increases with the number of modelled disorders which are used as dimensions, we span the CFPS within the 216 classes available from DeepGestalt. A low cophenetic distance between individuals #1 and #2 was achieved, with the centroid of Hutchison-Gilford Progeria being the closest of all previously modelled gene-phenotypes.

FaceNet analysis

We used a pre-trained FaceNet model (version 20170512-110547) producing a 128-dimensional embedding vector from a facial picture on a dataset of individual photos.

For face verification the embeddings are created as follows: in the case of three photo instances, a photo A of an individual, a different photo A' of the same individual, and a photo B of another individual, the distance between the embedding vectors A and A' is going to be minimized and the distance between the vectors A and B maximized. This is achieved through triplet-loss network training where the triplets are A, A' and B which are the target, the positive and the negative example respectively. We hypothesize that the distance between individuals with the same disorder (A1,...,An) is going to be smaller than the distances between individuals with different disorders (B1,...,Bn).

Our dataset included frontal photos of 265 individuals, amongst them images from 22 patients with a mutation in the *LMNA* gene, 11 with a mutation in the *PYCR1* gene and photos of the two individuals with a mutation in the *LEMD2* gene. In order to include all of the visual information we have about the two individuals (3 pictures per individual from 3 different ages), we averaged out their three respective vectors into one. Based on the representation vectors created by the model, we calculated the pairwise Euclidean distances between all individuals. The distances were then hierarchically clustered, producing a dendrogram shown in Fig S2B. In this novel comparison space, the two individuals fall right next to each other.

Supplemental References

1. Kuczmarski RJ, Ogden CL, Guo SS, Grummer-Strawn LM, Flegal KM, Mei Z, Wei R, Curtin LR, Roche AF, Johnson CL. (2002). 2000 CDC Growth Charts for the United States: methods and development. Vital Health Stat. 11(246), 1-190.

2. Kromeyer-Hauschild K, Wabitsch M, Kunze D, Geller F, Geiß HC, Hesse V, von Hippel A, Jaeger U, Johnsen D, Korte W, et al. (2001). Perzentile für den Body-mass-Index für das Kindes- und Jugendalter unter Heranziehung verschiedener deutscher Stichproben. Monatsschr Kinderheilkd 149, 807.

3. Kong A, Frigge ML, Masson G, Besenbacher S, Sulem P, Magnusson G, Gudjonsson SA, Sigurdsson A, Jonasdottir A, Jonasdottir A, et al. (2012). Rate of de novo mutations and the importance of father's age to disease risk. Nature. 488(7412):471-5.

4. Conrad DF, Keebler JE, DePristo MA, Lindsay SJ, Zhang Y, Casals F, Idaghdour Y, Hartl CL, Torroja C, Garimella KV, et al. (2011). Variation in genome-wide mutation rates within and between human families. Nat Genet. 43(7):712-4.

5. Sun JX, Helgason A, Masson G, Ebenesersdóttir SS, Li H, Mallick S, Gnerre S, Patterson N, Kong A, Reich D, et al. (2012). A direct characterization of human mutation based on microsatellites. Nat Genet. 44(10):1161-5.

6. Michaelson JJ, Shi Y, Gujral M, Zheng H, Malhotra D, Jin X, Jian M, Liu G, Greer D, Bhandari A, et al. (2012). Whole-genome sequencing in autism identifies hot spots for de novo germline mutation. Cell. 151(7):1431-42.

7. Besenbacher S, Sulem P, Helgason A, Helgason H, Kristjansson H, Jonasdottir A, Jonasdottir A, Magnusson OT, Thorsteinsdottir U, Masson G, et al. (2016). Multi-nucleotide de novo Mutations in Humans. PLoS Genet. 12(11):e1006315.

8. Altmüller J, Motameny S, Becker C, Thiele H, Chatterjee S, Wollnik B, Nürnberg P. (2016). A systematic comparison of two new releases of exome sequencing products: the aim of use determines the choice of product. Biol Chem. 397(8), 791-801.

9. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB et al. (2016). Exome Aggregation Consortium. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 536(7616), 285-91.

10. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, et al. (2015) ACMG Laboratory Quality Assurance Committee. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. 17(5), 405-24.

11. Stenson PD, Ball EV, Mort M, Phillips AD, Shiel JA, Thomas NS, Abeysinghe S, Krawczak M, Cooper DN. (2003). Human Gene Mutation Database (HGMD): 2003 update. Hum Mutat. 6, 577-81.

12. Milborrow, Stephen and Fred Nicolls. (2014). "Active shape models with SIFT descriptors and MARS." 2014 International Conference on Computer Vision Theory and Applications (VISAPP) 2, 380-387.

13. Ferry Q, Steinberg J, Webber C, FitzPatrick DR, Ponting CP, Zisserman A, Nellåker C. (2014). Diagnostically relevant facial gestalt information from ordinary photos. Elife. 3:e02020.