

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

	Page
Supplemental Note #1: Case Reports	2
Supplemental Note #2: Comment on <i>de novo</i> SNVs	5
Supplemental Figures	6
Supplemental Materials and Methods	9
Exome sequencing	9
Molecular biology	11
Cell lines	11
Immunofluorescence	11
Antibodies	12
Expression constructs	12
Western blotting	12
Facial analysis	13
Creation of composite images	13
DeepGestalt analysis	13
FaceNet analysis	14
Supplemental References	15

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 generation³⁻⁷, 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 data⁷.

Figure S1

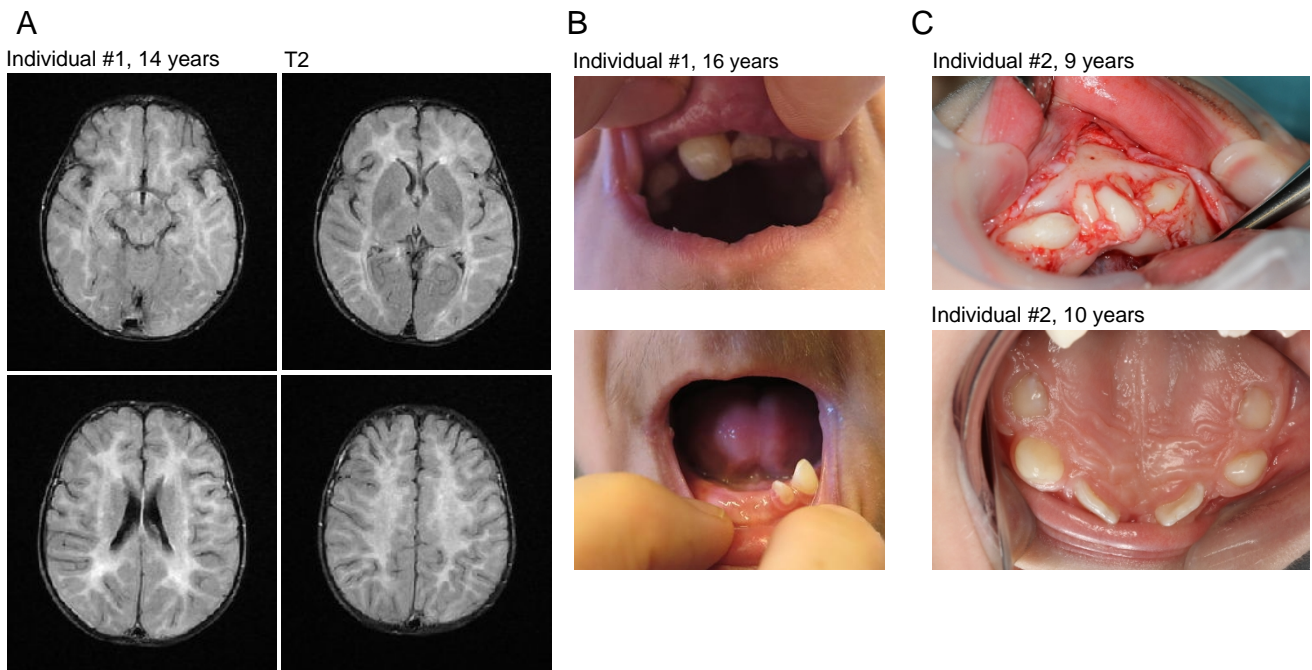


Figure S1

Cranial MRI and dental anomalies.

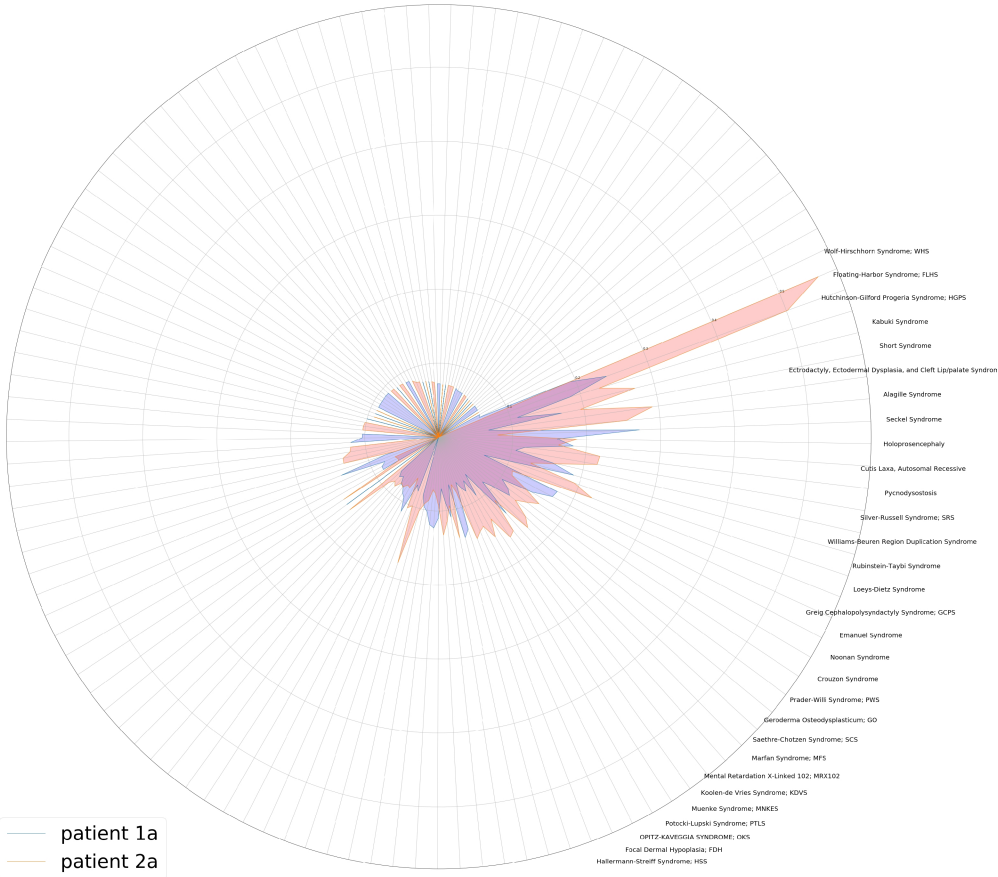
A: Cranial MRI of individual #1 at the age of 14 years shows widespread diffuse hyperintensities of the cerebral white matter in T2 and T2 FLAIR across all cerebral cross-sections, indicating hypomyelination.

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



B

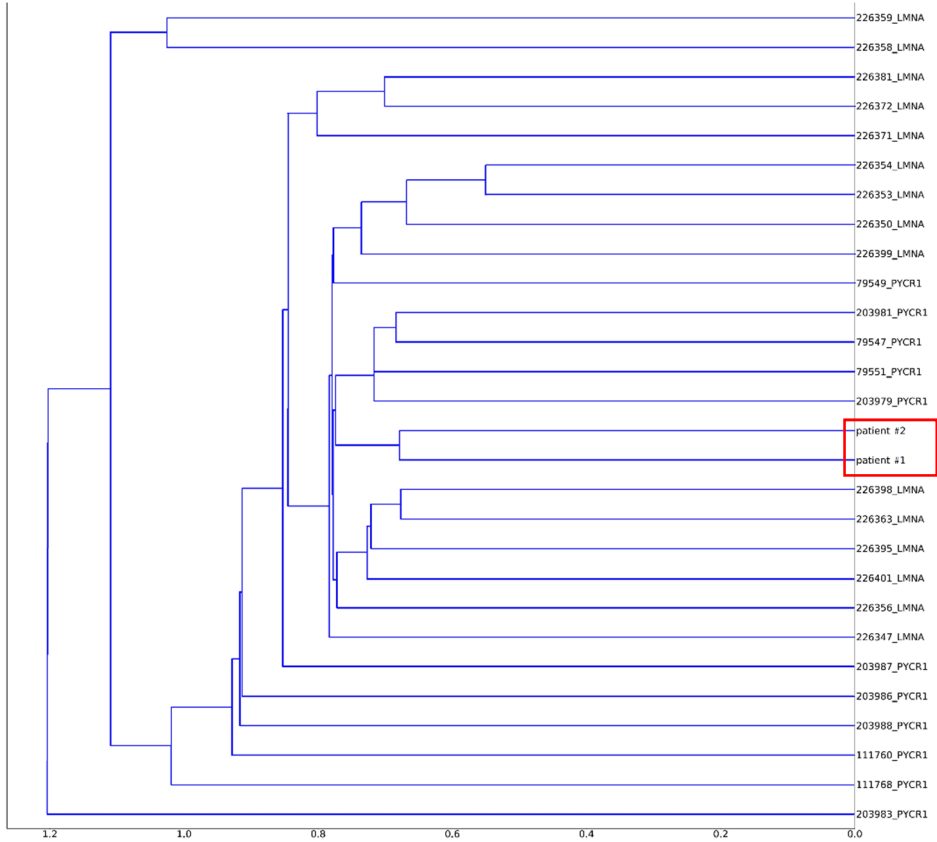


Figure S2

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 described⁸.

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 database¹¹. 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% CO₂ 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 TTACTIONTATCGTCGTCATCCTTGTAATCTCGCTCTGAGTCAGAGAAGG (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 library¹² 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 Hutchinson-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, Abeyasinghe 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.