ARTICLE

Evaluation of Face2Gene using facial images of patients with congenital dysmorphic syndromes recruited in Japan

Hiroyuki Mishima [1] · Hisato Suzuki [2] · Michiko Doi[3] · Mutsuko Miyazaki[4] · Satoshi Watanabe[5] ·
Tadashi Matsumoto[6] · Kanako Morifuji[7] · Hiroyuki Moriuchi[5] · Koh-ichiro Yoshiura[1] · Tatsuro Kondoh[6] ·
Kenjiro Kosaki[2]

Abstract

An increasing number of genetic syndromes present a challenge to clinical geneticists. A deep learning-based diagnosis assistance system, Face2Gene, utilizes the aggregation of "gestalt," comprising data summarizing features of patients' facial images, to suggest candidate syndromes. Because Face2Gene's results may be affected by ethnicity and age at which training facial images were taken, the system performance for patients in Japan is still unclear. Here, we present an evaluation of Face2Gene using the following two patient groups recruited in Japan: Group 1 consisting of 74 patients with 47 congenital dysmorphic syndromes, and Group 2 consisting of 34 patients with Down syndrome. In Group 1, facial recognition failed for 4 of 74 patients, while 13–21 of 70 patients had a diagnosis for which Face2Gene had not been trained. Omitting these 21 patients, for 85.7% (42/49) of the remainder, the correct syndrome was identified within the top 10 suggested list. In Group 2, for the youngest facial images taken for each of the 34 patients, Down syndrome was successfully identified as the highest-ranking condition using images taken from newborns to those aged 25 years. For the oldest facial images taken at ≥20 years in each of 17 applicable patients, Down syndrome was successfully identified as the highest- and second-highest-ranking condition in 82.2% (14/17) and 100% (17/17) of the patients using images taken from 20 to 40 years. These results suggest that Face2Gene in its current format is already useful in suggesting candidate syndromes to clinical geneticists, using patients with congenital dysmorphic syndromes in Japan.

✉ Hiroyuki Mishima
hmishima@nagasaki-u.ac.jp

1 Department of Human Genetics, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, Japan

2 Center for Medical Genetics, Keio University School of Medicine, Tokyo, Japan

3 Department of Pediatrics, Nagasaki University Hospital, Nagasaki, Japan

4 Department of Pediatrics, Nagasaki Prefectural Children Medical Welfare Center, Isahaya, Japan

5 Department of Pediatrics, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, Japan

6 Division of Developmental Disabilities, Misakaenosono Mutsumi Developmental, Medical and Welfare Center, Isahaya, Japan

7 Department of Nursing, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, Japan

Introduction

Craniofacial phenotypes are highly informative for the diagnosis of congenital genetic diseases. Before and even after genetic tests including next-generation sequencing-based ones, it has become an increasing challenge for clinical geneticists to make a definitive diagnosis in view of the increasing number of identified genetic syndromes (at over 8000) and their expanding range of atypical phenotypes [1]. Deep phenotyping and the integration of all biomedical knowledge are expected to assist in decision-making among clinical geneticists. In particular, phenotypes described in standardized and structured terminologies, such as the Human Phenotype Ontology [2], have been demonstrated to effectively prioritize candidate syndromes in applications including Phenomizer [3] and Pub-CaseFinder [4]. However, a comprehensive description of craniofacial phenotypes is difficult even for experienced clinical geneticists because their recognition of facial phenotypes is sometimes subjective, and the sum of

descriptions of partial facial features is not always representative of the entire facial phenotype. Thus, the introduction of computational solutions to analyze facial images directly is long-awaited in the field of clinical genetics. The facial dysmorphology novel analysis technology (FDNA), developed by FDNA Inc. (Boston, MA, USA), is a solution to this that works by combining computational facial recognition and a clinical knowledge database [5], which is implemented in the Face2Gene web application (https://www.face2gene.com/). Its facial recognition framework DeepGestalt [6] recognizes two-dimensional facial images to detect facial landmarks, followed by facial subregion detection, subregion feature extraction using a deep convolutional neural network, and feature integration to build numeric information called a "gestalt." Although gestalt data represent all of the facial features of an image, they do not contain any personal identifiable information. Gestalt information aggregation and comparison of patients with definitively diagnosed genetic syndromes enable ranking of candidate syndromes for diagnosis.

In previous studies, Face2Gene was used to evaluate the craniofacial phenotypes of syndromes, including Cornelia de Lange syndrome [7], fetal alcohol spectrum disorders [8], inborn errors of metabolism [9], SATB2-associated syndrome [10], Emanuel syndrome and Pallister–Killian syndrome [11], Ayme–Gripp syndrome [12], glycosylphosphatidylinositol biosynthesis defect [13], and phosphomannomutase-2 deficiency congenital disorder of glycosylation (PMM2-CDG) [14]. The performance of an application of a machine learning-based system depends on the quality and quantity of training datasets. Because the majority of clinical information aggregated in Face2Gene is on patients of European descent, the performance of this system for other populations remains unclear. Gurovich of FDNA Inc. and colleagues reported that Face2Gene successfully suggested the correct syndrome in 90.5% without mentioning the ethnic background of the cases and subjects in the training dataset [6]. Moreover, a study evaluating Face2Gene using patients with Down syndrome (DS) [15] reported sensitivity rates of Face2Gene in suggesting DS within the top 10 syndromes in patients recruited in Europe and Africa of 80% (16/20) and 36.8% (7/19), respectively. Moreover, the latest report on patients with DS in Thailand [16] described sensitivity of the highest-ranked suggested syndrome and the syndrome being within the top 10 suggested conditions of 90% (27/30) and 100% (30/30), respectively.

In this study, we aimed to evaluate the performance of Face2Gene in analyzing the facial images of patients recruited in Japan as a tool for clinical geneticists to prioritize the diagnosis of candidate genetic syndromes. Towards this aim, we first analyzed patients with diverse congenital dysmorphic syndromes. Second, we also evaluated how facial images taken at a young age can be applied for the ranking of candidate syndromes, by analyzing facial images of patients with DS acquired at multiple ages. The evaluation was undertaken in Face2Gene from late 2018 to early 2019.

Materials and methods

Participants and ethical issues

In this study, we analyzed the following two case groups: Group 1, patients with diverse congenital dysmorphic syndromes (Supplemental Table 1); and Group 2, patients with DS (Supplemental Table 2). All patients had undergone medical treatment at Misakaenosono Mutsumi Developmental, Medical and Welfare Center, Isahaya, Nagasaki, Japan (Mutsumi-no-ie). Patients in Group 1 had definitive diagnoses based on genetic tests and/or clinical diagnoses based on their phenotypes. Group 1 consisted of 74 patients with 47 different syndromes. All of the 34 patients in Group 2 had a definitive diagnosis based on G-banding. In most cases in Group 1 and all cases in Group 2, written informed consent was obtained from the participants, that is, the patients and/or their families. The opt-out consent model was applied to some of the Group 1 cases. This study was approved by institutional review boards of Mutsumi-no-ie, Nagasaki University Hospital, and Keio University. In Group 1, most of the facial images were taken by healthcare professionals, except in four cases of Sotos syndrome for which photographs were taken by family members. In Group 2, facial images of the patients with DS were taken at multiple ages for individual patients by family members. Although we did not collect information on the ancestry of the patients, we assumed that both groups consisted of people of Japanese ethnicity. The images used in this study were not normalized; the quality of the images varied in terms of resolution, brightness, and saturation. There were no obvious differences in quality between the images taken by the healthcare professionals and by the patients' families.

Information processing

The Face2Gene system software was provided by FDNA Inc. and installed on a server at the Center for Medical Genetics, Keio University School of Medicine, Tokyo, Japan (Keio server) (Fig. 1). The Keio server was connected to the internet and protected by a network firewall. Facial gestalt machine learning data were synchronized periodically with Face2Gene central servers on a secure cloud computing environment administered by FDNA (central server). For Group 2, multiple facial images of the same individuals were uploaded to assess those images taken at
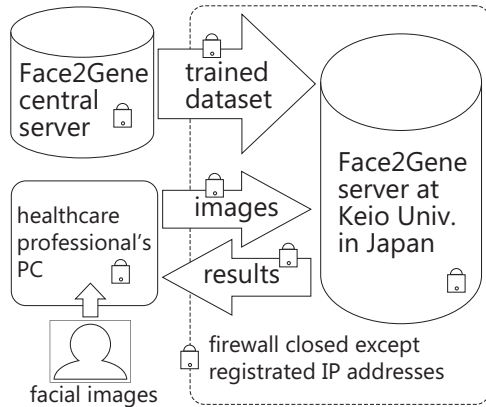
the youngest age that enabled a successful diagnosis of DS. Face2Gene results in this study were obtained from September 2018 to March 2019. Cases in this study had not previously been submitted to Face2Gene. To assess whether Face2Gene had already been trained for the syndromes of each of the patients in this study, we assessed composite photo availability by searching the name of the syndrome in the "refine phenotype" window. For each suggested syndrome, Face2Gene shows a barplot called the "gestalt

Fig. 1 Information flow in this study. Lock signs indicate information protection in data storage (encryption), data transmission via the internet (encryption), and access control (firewall and ID/password)

level," which indicates levels of "high," "medium," and "low." We used this information to evaluate the gestalt similarity of the analyzed patient and the trained syndrome.
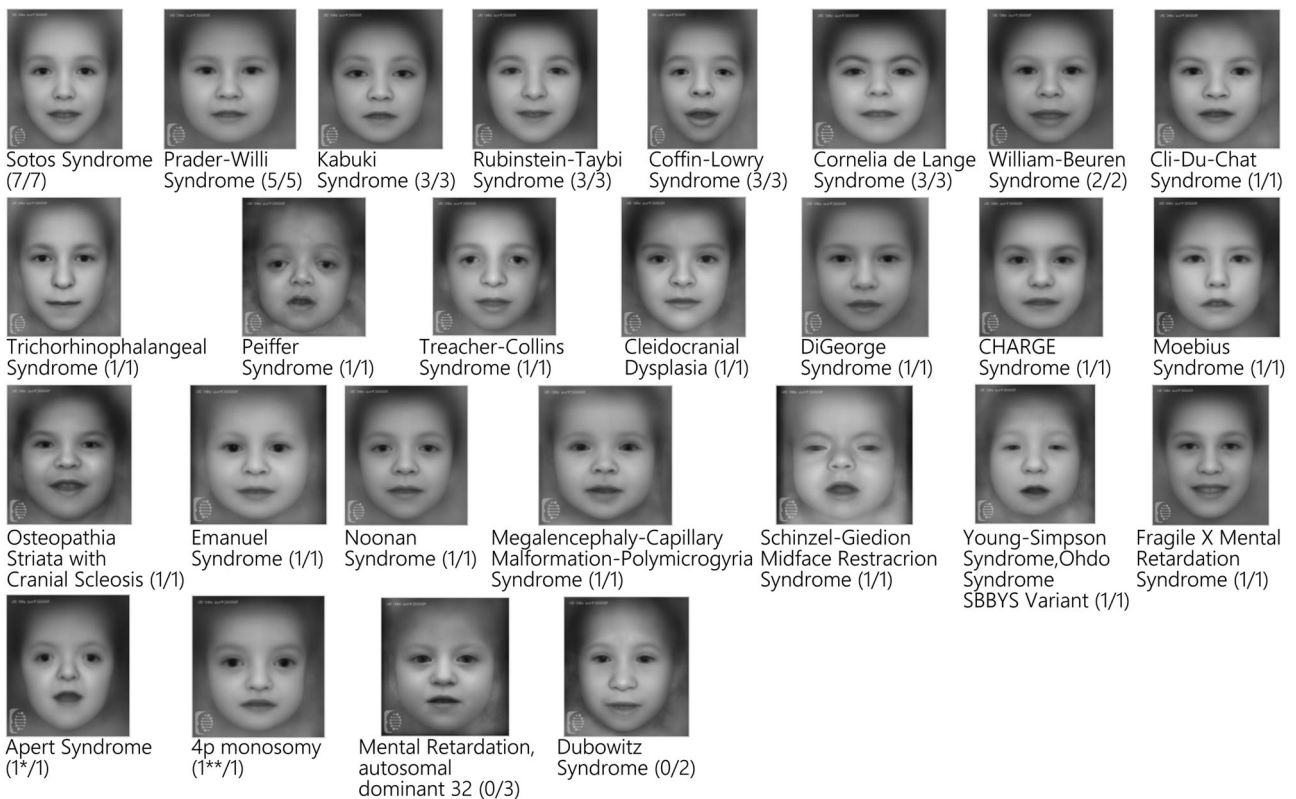
Results

Face2Gene evaluation for rare genetic diseases in a population in Japan

In accordance with a previous report [6], we evaluated the performance of Face2Gene by the sensitivity in terms of placement of the correct syndrome within the top K conditions. For instance, top 10 sensitivity means that the correctly diagnosed syndrome was suggested as one of the top 10 in the sorted suggestion list of Face2Gene. We adopted $K = 1$, 3, and 10 because we assumed that these values are reasonable in situations in which clinical geneticists prioritize candidate syndromes for diagnosis. First, to evaluate the present performance of Face2Gene to correctly suggest the status of patients with diverse congenital dysmorphic syndromes in Japan, we assessed Group 1 patients (Supplemental Table 1). Available Face2Gene composite facial images of clinically or definitively diagnosed syndromes of Group 1 are shown in Fig. 2. Note that

Fig. 2 Available composite facial images of Group 1 syndromes generated by Face2Gene using any of the patients' images. Numbers of concordant matches ranked within the top 10 and total number of concordant matches ranked within the top 10 and total number of cases in Group 1 are also denoted. *Suggested as Crouzon syndrome in the fourth-ranked position. **Suggested as Wolf–Hirschhorn syndrome in the third-ranked position

Table 1 Top K sensitivity of Face2Gene

Case group	Top 1 sensitivity[a]	Top 3 sensitivity[a]	Top 10 sensitivity[a]
Group 1: all cases	42.9% (30/70)	52.9% (37/70)	60.0% (42/70)
Group 1: Face2Gene-trained syndromes	61.2% (30/49)	75.5% (37/49)	85.7% (42/49)
Group 2: cases with DS using youngest facial images	100% (34/34)	100% (34/34)	100% (34/34)
Group 2: cases with DS using oldest facial images taken at ≥20 years	82.2% (14/17)	100% (17/17)	100% (17/17)

[a]Only images with successful face recognition were counted. Numbers in parentheses indicate numbers of cases.

these composite images were not generated using any of the patients' images. Recognition of the patient's face itself, which was enabled by facial landmark recognition, was successful in 94.6% of cases (70 of 74 cases) in Group 1. Images of four failed cases were those of newborn babies, the faces of whom were partially covered by tubes or taken at low resolution. Top 1 sensitivity was achieved in 42.9% of cases (30 of 70 cases). Top 3 sensitivity and top 10 sensitivity were achieved in 52.9% (37 of 70 cases) and 60.0% (42 of 70 cases), respectively. Distinguished by Face2Gene composite photo availability, Face2Gene had not been trained for the syndromes in 21 of the 70 cases. Omitting these 21 cases, the top 1, top 3, and top 10 sensitivity rates were 61.2% (30/49), 75.5% (37/49), and 85.7% (42/49), respectively. Among 30 cases in which first-ranked concordant results were obtained, 19 cases obtained the "high" first-ranked gestalt level, while those of the second-ranked suggestion were "med" or "low." In Group 1, a patient with Apert syndrome was suggested to be a patient with Crouzon syndrome, as the fourth rank in the suggestion list. In this case, we considered that the suggestion was accurate within the top 10 because these craniosynostosis syndromes share the same causal gene, *FGFR2*. In two cases with Dubowitz syndrome and three cases with mental retardation, autosomal dominant 32 (MRD32), concordant results could not be obtained despite high-quality images. We analyzed facial images in the recent literature about Dubowitz syndrome [17] and MRD32 [18]. For one of two facial images of patients with Dubowitz syndrome, concordant results were successfully obtained in the first rank. Moreover, for two of four images of patients with MRD32, concordant results were obtained at the 21st and 23rd ranks.

Face2Gene evaluation for facial images of Japanese patients with DS at multiple ages

To evaluate the effect of age at which facial images are taken on Face2Gene performance, we focused on Group 2, consisting of the youngest facial images (Supplemental Table 2) and oldest facial images taken at ≥20 years (Supplemental Table 3). For the youngest facial images, the top

1 sensitivity rate was 100% in all 34 cases with DS, when using facial images from at least the second youngest age (Table 1). The youngest ages at which facial images were obtained and produced first-rank results for DS spanned newborn to 25 years (mean: 3 years 11 months, median: 1 year 2 months). In 26 of the 34 cases, the gestalt levels of the first-ranked suggestion of DS were "high," while those of the second-ranked suggestion were "med" or "low." There was a failure to obtain first-ranked DS results using the youngest facial images in three different cases (3 weeks, 2 years, and 12 years). Two of three of these images had a low resolution, and one of them showed a face partially hidden by eyeglasses. For the oldest facial images taken at ≥20 years, spanning from 20 to 40 years (mean: 26 years, median: 25 years), the top 1 and 3 sensitivity rates of the 17 cases were 82.2% (14/17) and 100% (17/17), respectively. In 14 of these 17 cases, the gestalt levels of the first-ranked suggestion of DS were "high".

Discussion

In this study, we evaluated the performance of Face2Gene in analyzing the facial images of patients recruited in Japan and found that this tool already has clinical utility despite its training dataset being assumed not to include enough cases from a concordant population. Because Face2Gene Deep-Gestalt itself does not perform differential diagnosis automatically and always shows the top 30 matches, we did not discuss its specificity in this study. Compared with a previous study that reported that the overall top 10 sensitivity rate was 90.6% [6], our results from Group 1 may be explained by the difference in ethnicity and insufficient training information in extremely rare genetic syndromes. In Group 1, for two patients with Dubowitz syndrome and MRD32, these conditions were not listed within the top 10 suggestions, despite relatively high image quality. Owing to the lack of access to information about the Face2Gene internal training data set, this study has a limitation regarding our ability to discuss the quality and quantity of the dataset. Nonetheless, the dataset was expected to at least contain information on both syndromes

because composite face images of both syndromes are available (Fig. 2). The results for MRD32 may be explained by an insufficient number of total cases and/or of cases with similar ethnicity. Because the molecular background of Dubowitz syndrome is still unknown, the results for the syndrome may also be explained by its heterogeneity, that is, patients with Dubowitz syndrome in the Face2Gene training dataset and in this study have different molecular backgrounds. In contrast to the findings in Group 1, the sensitivity of Group 2 with DS was comparable to that in previous studies [15, 16]. This may be explained by training with a sufficient number of DS cases already having been performed and/or that the gestalt of DS was sufficiently informative to overcome the effects of ethnic background. The analysis of each youngest facial image in Group 2 showed that facial images of newborn babies are sufficient to achieve a match to DS at the highest rank in the list of suggested syndromes. By contrast, the analysis of each oldest facial image taken at ≥20 years showed slightly lower top 1 sensitivity. The ages at which photographs are taken have been reported to affect Face2Gene results [9]. Although this study has limitations in terms of the sample size and the quality control of facial images, our results may suggest that the effect of the ages at which facial images are taken is stronger at older ages, at least in cases with DS.

Face2Gene is suitable for use even by healthcare professionals with little experience in terms of both computer science and clinical genetics. Using Face2Gene without consultation with clinical geneticists may, however, cause misevaluation of the results, including considering them as a definitive diagnosis. Consultation with clinical geneticists is essential, including for genetic tests such as cytogenetics, Sanger sequencing, microarray analysis, whole-exome and whole-genome sequencing, and related bioinformatics. Clinical geneticists also play an important role in offering medical treatment and social assistance for patients and their families. Advertisements about the availability of computer assistance in clinics including Face2Gene for patients, their families, and medical professionals are required to maximize its advantages and minimize misevaluation.

This study suggested the usefulness of Face2Gene in ranking candidate syndromes for clinical geneticists' diagnoses of patients recruited in Japan. Ongoing case registration and aggregation in Face2Gene for Japanese patients with congenital dysmorphic syndromes is necessary to improve its performance, not only for patients in Japan but also for those from other populations.

Compliance with ethical standards

References

1. Kikuiri T, Mishima H, Imura H, Suzuki S, Matsuzawa Y, Nakamura T, et al. Patients with SATB2-associated syndrome exhibiting multiple odontomas. Am J Med Genet A. 2018;176A: 2614–22.

2. Groza T, Köhler S, Moldenhauer D, Vasilevsky N, Baynam G, Zemojtel T, et al. The human phenotype ontology: semantic unification of common and rare disease. Am J Hum Genet. 2015;97:111–24.

3. Smedley D, Jacobsen JOB, Jäger M, Köhler S, Holtgrewe M, Schubach M, et al. Next-generation diagnostics and disease-gene discovery with the Exomiser. Nat Protoc. 2015;10:2004–15.

4. Fujiwara T, Yamamoto Y, Kim J-D, Buske O, Takagi T. PubCaseFinder: a case-report-based, phenotype-driven differential-diagnosis system for rare diseases. Am J Hum Genet. 2018;103: 389–99.

5. Gripp KW, Baker L, Telegrafi A, Monaghan KG. The role of objective facial analysis using FDNA in making diagnoses following whole exome analysis. Report of two patients with mutations in the BAF complex genes. Am J Med Genet A. 2016;170:1754–62.

6. Gurovich Y, Hanani Y, Bar O, Nadav G, Fleischer N, Gelbman D, et al. Identifying facial phenotypes of genetic disorders using deep learning. Nat Med. 2019;25:60.

7. Basel-Vanagaite L, Wolf L, Orin M, Larizza L, Gervasini C, Krantz ID, et al. Recognition of the Cornelia de Lange syndrome phenotype with facial dysmorphology novel analysis. Clin Genet. 2016;89:557–63.

8. Valentine M, Bihm DCJ, Wolf L, Hoyme HE, May PA, Buckley D, et al. Computer-aided recognition of facial attributes for fetal alcohol spectrum disorders. Pediatrics. 2017;140: e20162028.

9. Pantel JT, Zhao M, Mensah MA, Hajjir N, Hsieh T-C, Hanani Y, et al. Advances in computer-assisted syndrome recognition by the example of inborn errors of metabolism. J Inherit Metab Dis. 2018;41:533–9.

10. Zarate YA, Smith-Hicks CL, Greene C, Abbott M-A, Siu VM, Calhoun ARUL. et al. Natural history and genotype-phenotype correlations in 72 individuals with SATB2-associated syndrome. Am J Med Genet A. 2018;176:925–35.

11. Liehr T, Acquarola N, Pyle K, St-Pierre S, Rinholm M, Bar O, et al. Next generation phenotyping in Emanuel and Pallister–Killian syndrome using computer-aided facial dysmorphology analysis of 2D photos. Clin Genet. 2018;93:378–81.

12. Amudhavalli SM, Hanson R, Angle B, Bontempo K, Gripp KW. Further delineation of Aymé-Gripp syndrome and use of automated facial analysis tool. Am J Med Genet A. 2018;176: 1648–56.

13. Knaus A, Pantel JT, Pendziwiat M, Hajjir N, Zhao M, Hsieh T-C, et al. Characterization of glycosylphosphatidylinositol biosynthesis defects by clinical features, flow cytometry, and automated image analysis. Genome Med. 2018;10:3.

14. Martinez-Monseny A, Cuadras D, Bolasell M, Muchart J, Arjona C, Borregan M, et al. From gestalt to gene: early predictive dysmorphic features of PMM2-CDG. J Med Genet. 2018;56:236–45. Epub ahead of print

15. Lumaka A, Cosemans N, Lulebo Mampasi A, Mubungu G, Mvuama N, Lubala T, et al. Facial dysmorphism is influenced by ethnic background of the patient and of the evaluator. Clin Genet. 2017;92:166–71.

16. Vorravanpreecha N, Lertboonnum T, Rodjanadit R, Sriplienchan P, Rojnueangnit K. Studying down syndrome recognition probabilities in Thai children with de-identified computer-aided facial analysis. Am J Med Genet A. 2018;176:1935–40.

17. Stewart DR, Pemov A, Johnston JJ, Sapp JC, Yeager M, He J, et al. Dubowitz syndrome is a complex comprised of multiple, genetically distinct and phenotypically overlapping disorders. PLOS One. 2014;9:e98686.

18. Millan F, Cho MT, Retterer K, Monaghan KG, Bai R, Vitazka P, et al. Whole exome sequencing reveals de novo pathogenic variants in KAT6A as a cause of a neurodevelopmental disorder. Am J Med Genet A. 2016;170:1791–8.