GeNeXpreSS

Next-Generation Phenotyping in the Next-Generation Sequencing Era

Somya Srivastava

Department of Medical Genetics, Sanjay Gandhi Post Graduate Institute of Medical Sciences, Lucknow, India

Correspondence to: Dr Somya Srivastava *Email:* somyasrivastava18@gmail.com

Next-generation phenotyping using computer vision algorithms in rare genomic neurodevelopmental disorders (van der Donk et al., 2019)

The authors used a novel algorithm by combining two computer algorithms: the Clinical Face Phenotype Space (CFPS) for facial dysmorphism and OpenFace for facial recognition. Using them, they detected the facial gestalt in three novel intellectual disability syndromes involving the genes *PACS1*, *PPM1D*, and *PHIP*. Significant facial similarity for all three syndromes was found. Hence information contained in the face can be used to delineate genetic entities including in novel ID syndromes with no previously known knowledge of a facial phenotype.

Evaluating Face2Gene as a tool to identify Cornelia de Lange syndrome by facial phenotypes (Latorre-Pellicer et al., 2020)

This study explored the sensitivity of artificial intelligence by means of Face2Gene technology for facial recognition in a group of 49 patients with molecularly confirmed Cornelia de Lange syndrome with mutations in *NIPBL*, *SMC1A*, *HDAC8* and *RAD21* genes. Cornelia de Lange, which can be diagnosed clinically but has features that vary widely in range and severity, was the first diagnosis in 41/49 patients and one of the top five diagnosis in 47/49 cases giving a sensitivity of 83.7% and 97.9% respectively. The other top five diagnoses were KBG syndrome, CHARGE syndrome, Rubinstein–Taybi syndrome and Moebius syndrome, with frequencies of 44.89% (22/49), 36.7% (18/49), 34.7% (17/49), and 18.4% (9/49), respectively. Although substantial difference in sensitivity regarding the age at which facial images were taken was not present, the sensitivity differed with the affected gene and presence of classical features with high sensitivity noted in patients with *NIBPL* variants (97%) and those with the classical phenotype (88.8%). Thus, each gene presented a different pattern recognition and this can be utilized for studying the genotype–phenotype correlations and to differentiate between genetic subtypes. For example, it has been described that thicker eyebrows are suggestive of a variation in *SMC1A* or *SMC3*, and females containing variants in *HDAC8* tend to have hypertelorism and a slightly bulbous nasal tip. Patients with *NIPBL* variants show pronounced facial features, compared to patients with *RAD21* variants who have less prominent features.

Computer-aided facial analysis in diagnosing dysmorphic syndromes in Indian children (Narayanan et al., 2019)

This study used Face2Gene to assess its utility in predicting the diagnosis in 51 Indian children with obvious facial dysmorphism and a definite molecular or cytogenetic diagnosis. A correct diagnosis as the first suggestion was found in 26 patients (50.9%) and as a part of the top ten suggestions was obtained in 37 patients (72.5%). This study highlights that the results of the software can change based on the ethnicity as the software was unable to provide a diagnosis in easily recognizable syndromes like Turner syndrome, Waardenburg syndrome and Wolf-Hirschhorn syndrome. Since Face2Gene learns from every solved case, its sensitivity is

likely to improve further with increasing use particularly in non-Caucasian populations.

PEDIA: prioritization of exome data by image analysis (Hsieh et al., 2019)

This paper assessed the value added by computer assisted image analysis (DeepGestalt) to the diagnostic yield on a cohort consisting of 679 individuals with 105 different monogenic disorders. For every case, scores from DeepGestalt were used along with the clinical features and CADD score of the causative variant and a PEDIA score was generated. The additional information from the photographs pushed the correct disease gene to the top 10 in 99% of all PEDIA cases from less than 45% when only CADD scores were used. The accuracy rate for the top one gene rose from 36–74% without DeepGestalt scores to 86–89% when artificial intelligence was used. The results were not affected by the ethnicity of the patients, however low accuracy was seen in very rare disease due to limited training for those particular genes.

Precision medicine integrating whole–genome sequencing, comprehensive metabolomics, and advanced imaging (Hou et al., 2019)

A cohort of 1190 adult volunteers underwent whole genome sequencing followed by deep phenotyping by metabolomics, advanced imaging, and clinical laboratory tests in addition to family/medical history. Integrating the results of WGS with deep phenotyping, 11.5% individuals had a pathogenic variant, thereby providing a plausible genetic cause for abnormal physiological measurements at the individual level of analysis. A high percentage of genotype phenotype correlation was observed for dyslipidemia, cardiomyopathy and arrhythmia, and diabetes and endocrine diseases. With deep phenotyping, heterozygous carriers of autosomal recessive diseases were also found to exhibit detectable phenotypic changes. Sixty-nine (5.8%) individuals had pathogenic/ likely pathogenic variants but did not have associated family history, medical history, or phenotypes detected in tests. This could be because of reduced penetrance, variable expressivity, or late onset of disease presentation.

References

1. Hsieh TC, et al. PEDIA: prioritization of exome data by image analysis. Genet Med 2019; 21: 2807–2814.
2. Hou YC, et al. Precision medicine integrating whole-genome sequencing, comprehensive metabolomics, and advanced imaging. Proc Natl Acad Sci USA 2020; 117: 3053–3062.
3. Latorre-Pellicer A, et al. Evaluating Face2Gene as a tool to identify Cornelia de Lange syndrome by facial phenotypes. Int J Mol Sci 2020; 21:1042.
4. Narayanan DL, et al. Computer-aided facial analysis in diagnosing dysmorphic syndromes in Indian children. Indian Pediatr 2019; 56: 1017–1019.
5. van der Donk R, et al. Next-generation phenotyping using computer vision algorithms in rare genomic neurodevelopmental disorders. Genet Med 2019; 21:1719–1725.