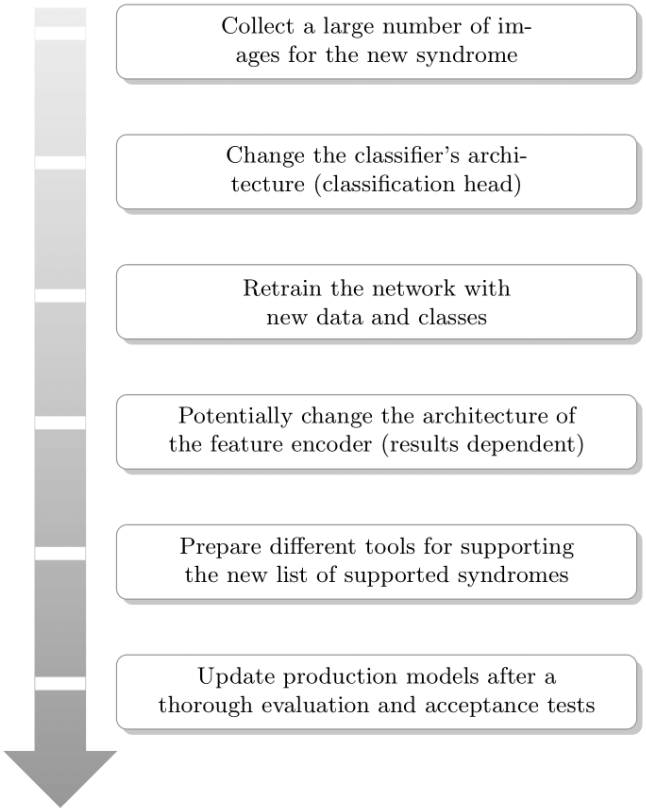
GestaltMatch: breaking the limits of rare disease matching using facial phenotypic descriptors

Supplementary Material

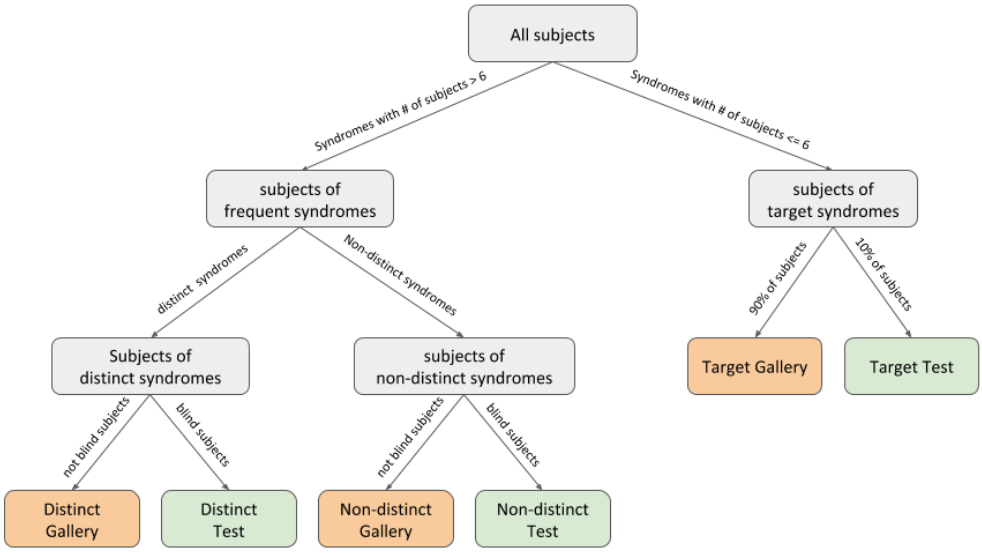
Hsieh, Tzung-Chien[1]; Bar-Haim, Aviram[2]; Nadav, Guy[2]; Pantel, Jean Tori[1,3]; Fleischer, Nicole[2]; Krawitz, Peter[1,*]

1. Institute of Genomic Statistics and Bioinformatics, University of Bonn, Bonn, Germany
2. FDNA Inc., Boston Massachusetts, United States
3. Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Institute of Medical Genetics and Human Genetics, Berlin, Germany

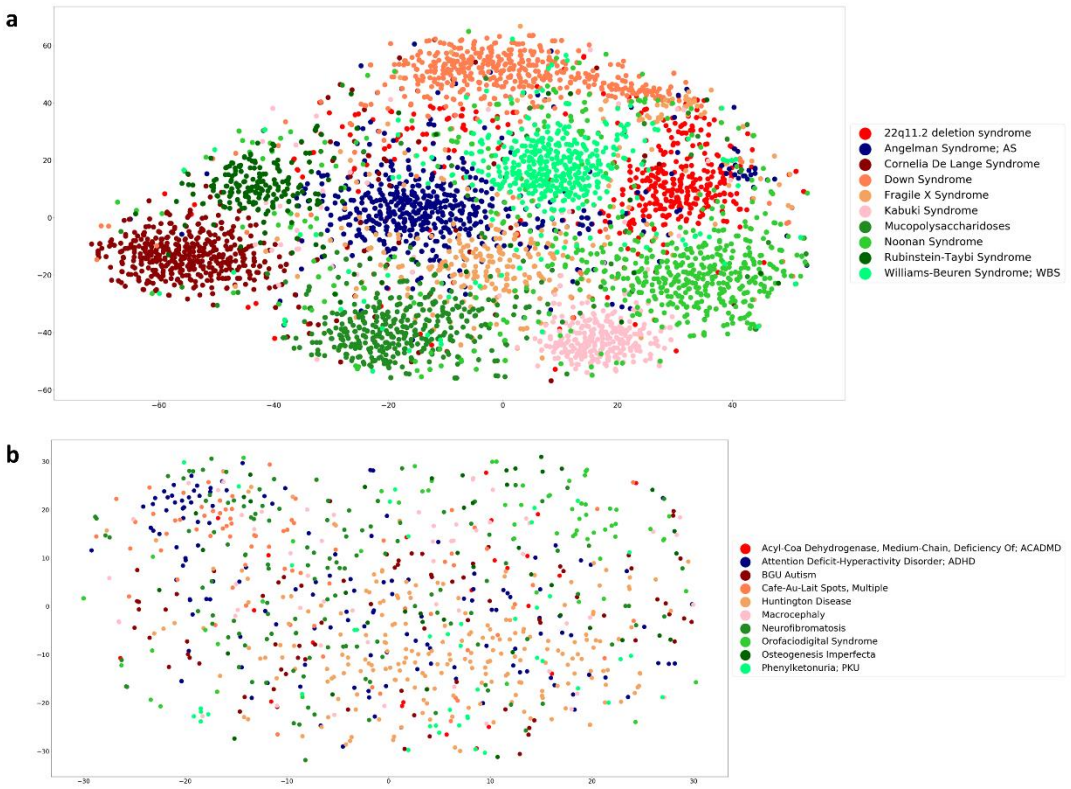
* corresponding author, pkrawitz@uni-bonn.de



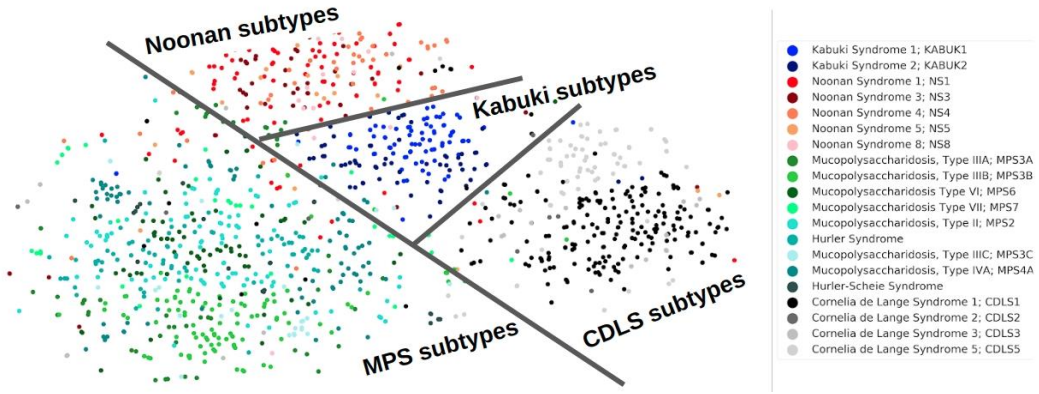
Supplementary Figure 1: The development flow for supporting new syndromes in DeepGestalt model. To include new syndromes into the "end-to-end" multi-syndromes classification framework such as DeepGestalt, the developer should go through these six steps. The model retraining might cost lots of money and time, which leads to the low scalability for supporting novel diseases or ultra-rare syndromes.



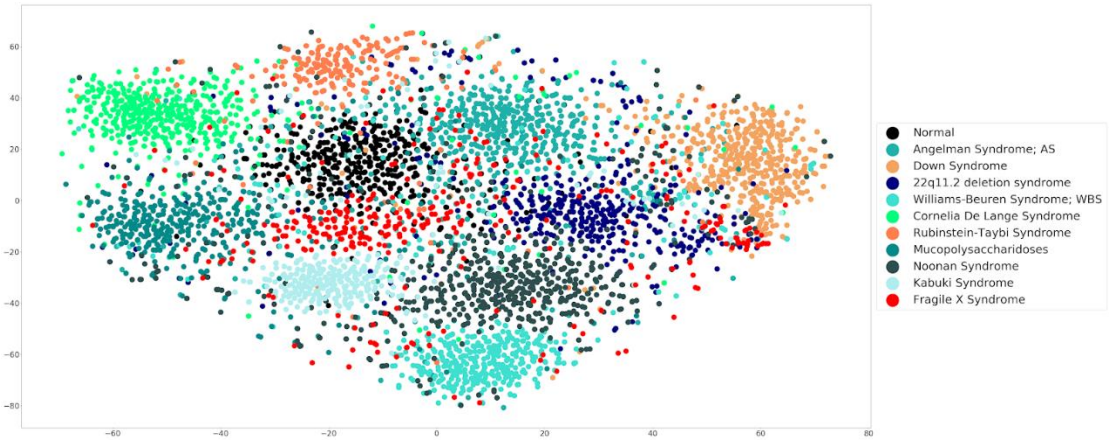
Supplementary Figure 2: Overview of data split in GestaltMatch. The data is first divided by the number of subjects in each syndrome. The syndromes with more than six subjects were assigned to frequent syndromes, and the syndromes with less than six subjects were assigned to rare syndromes. The rare syndromes was denoted as target syndromes. Frequent syndromes were further separated by whether DeepGestalt recognized the facial dysmorphic features of syndrome. If DeepGestalt does not recognize it, we denote it as non-distinct syndrome, and vice versa. Each category was further divided into the gallery and test set. We performed 10-fold cross-validation on target syndromes. For distinct and non-distinct syndromes, the subjects in the blind set, which is kept for validating DeepGestalt model training, were sampled in the test set.



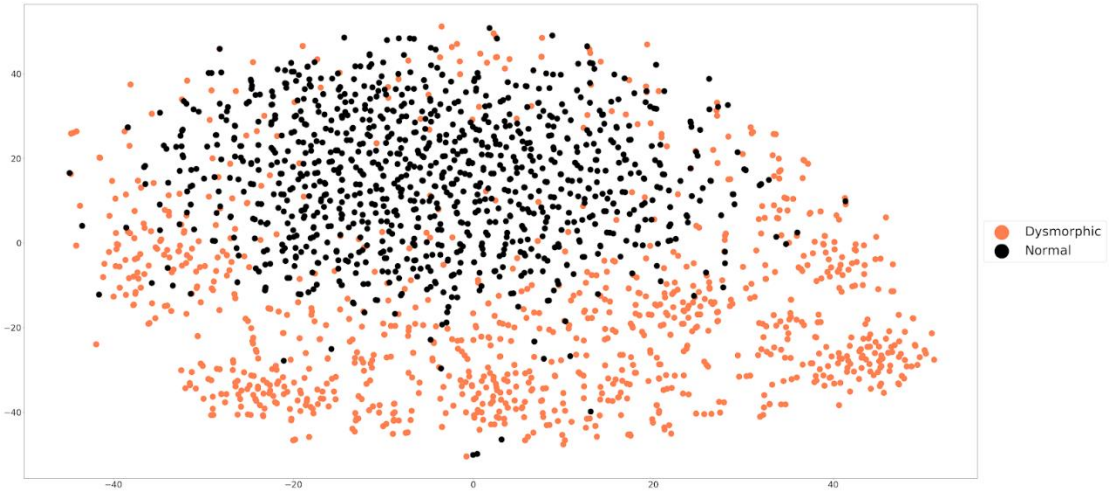
Supplementary Figure 3: t-SNE visualization of FPDs of ten distinct syndromes (a) and ten non-distinct syndromes (b).



Supplementary Figure 4: Hierarchical clustering of the phenotypic series of Kabuki syndrome, Noonan syndrome, Mucopolysaccharidosis, and Cornelia de Lange syndrome.



Supplementary Figure 5: Clustering of normal subjects and the ten syndromes with the most subjects.



Supplementary Figure 6: Clustering of normal subjects and the affected subjects.