Am. J. Hum. Genet. 77:999–1010, 2005

Discriminating Power of Localized Three-Dimensional Facial Morphology

Peter Hammond,[1] Tim J. Hutton,[1] Judith E. Allanson,[8] Bernard Buxton,[2] Linda E. Campbell,[4] Jill Clayton-Smith,[7] Dian Donnai,[7] Annette Karmiloff-Smith,[3] Kay Metcalfe,[7] Kieran C. Murphy,[6] Michael Patton,[5] Barbara Pober,[9] Katrina Prescott,[3] Pete Scambler,[3] Adam Shaw,[5] Ann C. M. Smith,[10,11] Angela F. Stevens,[4] I. Karen Temple,[12] Raoul Hennekam,[3] and May Tassabehji[7]

[1]Eastman Dental Institute, [2]Department of Computer Science, and [3]Institute of Child Health, University College London, [4]Institute of Psychiatry, King's College, and [5]St. Georges Hospital Medical School, London; [6]Royal College of Surgeons in Ireland, Dublin; [7]The Academic Unit of Medical Genetics and Regional Genetic Service, University of Manchester, St. Mary's Hospital, Manchester, United Kingdom; [8]Eastern Ontario Regional Genetics Program, Ottawa, Canada; [9]Children's Hospital, Boston; [10]National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; [11]Georgetown University Medical Center, Washington, DC; and [12]Wessex Clinical Genetics Service and Division of Human Genetics, Southampton University and Hospital Trust, Southampton, United Kingdom

Many genetic syndromes involve a facial gestalt that suggests a preliminary diagnosis to an experienced clinical geneticist even before a clinical examination and genotyping are undertaken. Previously, using visualization and pattern recognition, we showed that dense surface models (DSMs) of full face shape characterize facial dysmorphology in Noonan and in 22q11 deletion syndromes. In this much larger study of 696 individuals, we extend the use of DSMs of the full face to establish accurate discrimination between controls and individuals with Williams, Smith-Magenis, 22q11 deletion, or Noonan syndromes and between individuals with different syndromes in these groups. However, the full power of the DSM approach is demonstrated by the comparable discriminating abilities of localized facial features, such as periorbital, perinasal, and perioral patches, and the correlation of DSM-based predictions and molecular findings. This study demonstrates the potential of face shape models to assist clinical training through visualization, to support clinical diagnosis of affected individuals through pattern recognition, and to enable the objective comparison of individuals sharing other phenotypic or genotypic properties.

Introduction

Many genetic syndromes involve craniofacial abnormalities (Gorlin et al. 2001). Indeed, the facial gestalt often suggests a preliminary diagnosis before a clinical examination and genotyping are undertaken (Winter 1996). A single facial feature, such as nose shape, may even be sufficient to suggest a particular syndrome. Previously, anthropometric and two-dimensional (2D) photogrammetric studies have delineated craniofacial morphology in a variety of syndromes (e.g., Allanson et al. 1985, 1993; Sharland et al. 1993; Allanson and Cole 1996; Allanson and Hennekam 1997; Ward et al. 2000). Studies using 2D images have achieved an accuracy of 74% in intersyndrome discrimination comparing five syndromic groups, each with 6–13 individuals (Loos et al. 2003).

Recently, rapid, noninvasive three-dimensional (3D) imaging of the face surface has become available. The clinical usability of 3D images is considerable because the face is viewable from any angle and at closer proximity than most children, or even adults, would tolerate. Each image comprises a surface of ⩾20,000 points and 2D images of facial appearance. Unlike 2D images, 3D surfaces are robust to changes in illumination. It is possible to retrieve 3D data from a single 2D image, but this requires standard lighting conditions or a previously constructed lighting model.

3D face analysis using dense surface models (DSM) has proven successful in delineating facial morphology in Noonan syndrome (NS [MIM 163950]), 22q11 deletion syndrome (22q11DS [MIM 192430]), Bardet-Biedl syndrome (MIM 209900) (Beales et al. 1997), and Smith-Magenis syndrome (SMS [MIM 182290]), and in discriminating well between controls and individuals with NS and 22q11DS (Hammond et al. 2003*a*, 2003*b*, 2004). Full face surface analysis was employed in each of these studies.

Craniofacial anomalies are associated with each of the syndromes considered in this study. NS is a relatively frequently occurring syndrome (~1/2,000 live births) characterized by short stature, heart defects, webbed neck, chest deformity, mild learning disability, and characteristic facial features. The facial phenotype in NS

Received June 3, 2005; accepted for publication September 2, 2005; electronically published October 26, 2005.

Address for correspondence and reprints: Dr. Peter Hammond, Biomedical Informatics Unit, Eastman Dental Institute, University College London, 256 Grays Inn Road, London, WC1X 8LD, United Kingdom. E-mail: p.hammond@eastman.ucl.ac.uk

© 2005 by The American Society of Human Genetics. All rights reserved.
0002-9297/2005/7706-0011$15.00

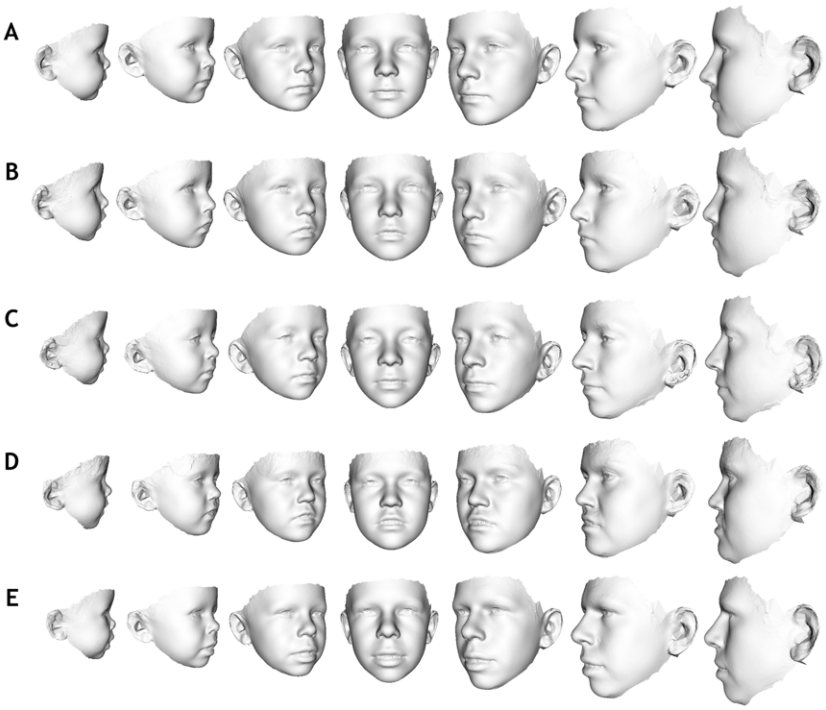

Figure 1 Facial phenotypes. *A*, Control; *B*, 22q11DS; *C*, NS; *D*, SMS; and *E*, WS. Each series is generated using a regression model of age and the first PCA mode of a DSM generated from faces in a homogeneous subgroup. The sequences do not use a common scale.

becomes less dysmorphic over time (Allanson et al. 1999). In ~40%–50% of patients, the genetic cause is a mutation in the protein tyrosine phosphatase PTPN11, located at 12q24.1 (Tartaglia et al. 2001). 22q11DS is thought to be the most frequently occurring microdeletion syndrome (1/4,000 live births), with a highly variable phenotype that includes palatal and cardiovascular anomalies, cognitive and behavioral abnormalities, and characteristic facial features (Scambler et al. 1992). Mouse models and rare Japanese mutations implicate haploinsufficiency for *TBX1* as the main genetic factor (Jerome and Papaioannou 2001; Yagi et al. 2003), although a role for neighboring genes has not been ruled out. SMS is rare (1/25,000 live births) and is usually caused by a microdeletion of ~4 Mb at 17p11.2 (Greenberg et al. 1991). The phenotype is characterized by infantile hypotonia, otolaryngological anomalies, expressive language delay, oral motor dysfunction, hearing loss, vision problems, short stature, small hands/feet (brachydactyly), scoliosis, developmental delay, expressed behavioral abnormalities that include major sleep disturbance, and facial dysmorphism. A limited number of patients have been found to harbor a mutation in *RAI1* instead of a microdeletion, suggesting that *RAI1* plays a significant role in the phenotype (Slager et al. 2003). Williams syndrome (WS [MIM 194050]), another neurodevelopmental disorder, occurs with a frequency of 1/20,000 live births. Affected individuals characteristically have an outgoing personality, cognitive abnormalities, specific cardiovascular defects, hypercalcaemia, and hyperacusis. It is usually caused by a chromosomal microdeletion of ~1.5 Mb at 7q11.23. Individuals with WS have a characteristic facies (Morris et al. 1988).

In the present study, we extended previous morphometric face analysis to SMS and WS and demonstrated analogous delineation and discrimination abilities, using DSM-based methods. Furthermore, we show that if we focus on smaller, more localized areas of the face, we can achieve discrimination rates comparable to that for the full face for all four syndromes. The discrimination analysis, in particular, identifies which areas of the face are more effective in distinguishing between controls and individuals with a particular syndrome and between syndromes. Finally, we demonstrate the full power of localized 3D facial morphology by demonstrating its correlation with the molecular analysis of four individuals with an initial, putative diagnosis of WS.

Subjects and Methods

Subject Recruitment and Data Capture

The controls, 185 individuals younger than 20 years and 132 20 years and older, were recruited as volunteers

or as unaffected siblings and parents. Individuals with a syndrome were recruited either at family support meetings and international conferences in Europe and the United States or through clinical collaborators. Before the study, all affected patients had been given a clinical diagnosis of one of the four syndromes (NS, 22q11DS, SMS, or WS). Patients were included in the study if they had been cytogenetically proven. They were excluded if judged by experienced dysmorphologists (J.E.A., R.H., K.M., M.P., A.C.M.S., and I.K.T.) to not have the facial features of the syndrome for which they had a putative diagnosis. Thus, the individuals studied comprised 696 controls and patients, all white.

The affected individuals consisted of 80 individuals with NS, 54 with SMS, 115 with 22q11DS, and 130 with WS. Of these, 16 patients with NS were known to have a PTPN11 mutation. Positive FISH tests were known for 46 of the SMS subset, 112 of the 22q11DS subset, and 109 of the WS subset.

The study was approved by the University College London Hospital Research Ethics Committee, and participants, parents, or guardians gave informed written consent (JREC 00/E042).

Facial Phenotypes

In patients with 22q11DS (fig. 1*b*), the face is somewhat subtle and not always obviously different from average. In general, one of the most striking features is increased nasal height, with relatively narrow nares and nasal base but with fullness above the tip of the nose. Eyes are often upslanted, with hooded lids. Ears may be cupped or unusual in shape. With increasing age, the exaggerated length of the nose is less obvious, but the nose remains unusual and somewhat tubular, with some broadening at the root and narrowing at the base. Fullness remains above and lateral to the tip.

The young child with NS (fig. 1*c*) has a tall forehead with some narrowing at the temples, wide-spaced and downslanting eyes, sometimes with droopiness of the lid, a short, broad nose, a well-grooved upper lip with full lips in a cupid's bow configuration, a small chin, and low-set, posteriorly angulated ears. Over time, the forehead becomes less dominant, while the chin becomes longer, leading to an inverted triangular facial shape. As the nose lengthens, it appears less broad. Flattening of the mid-face is probably apparent at all ages but may be more obvious in older individuals, when fullness of the cheeks is less prominent.

In patients with SMS (fig. 1*d*), the facial appearance can be quite subtle in early life but becomes more distinctive with age. One characteristic feature is the unusual shape of the upper lip, where the philtrum is "trapezoidal" and the philtral pillars appear thickened and everted. In profile, the eversion of the upper lip mirrors the protrusion of the tip of the nose. The face becomes increasingly square and "pugnacious" because of midface hypoplasia, especially in the older individual. The brow is heavy, accentuating deep, close-set eyes that are often upslanting. Synophrys is usually present. Over time, the jaw becomes broader and more prominent and prognathic relative to the upper face.

For WS (fig. 1*e*), in the young child, features include broad forehead, bitemporal narrowing, depressed nasal root, periorbital fullness, full nasal tip, flattening of the mid-face, long philtrum, full lips, wide mouth, and full cheeks with a small jaw. With age, in addition to vertical lengthening of the face, there is a more gaunt appearance, less fullness to the cheeks, a prominent supraorbital ridge, narrow and fairly prominent root of the nose, continuing flattening of the mid-face, more notable widening of the mouth, and full lips.

3D Image Capture and Preparation

3D face images were captured with commercial photogrammetric devices (3dMD and Surfim). Depending on the system used, the captured face surface contained 4,000–20,000 3D points. Each image was manually annotated with 21 3D landmarks: left and right endocanthion, exocanthion, palpebrale superius, palpebrale inferius, crista philtrum, cheilion, alare and lower lip third; nasion, pronasale, subnasale, labiale superius, and gnathion. For each of the syndrome-control and intersyndrome comparisons, 20 randomly generated 90%–10% training-test set pairs, stratified with respect to affected and unaffected totals, were generated from the appropriate set of landmarked face surfaces. Each training set was used to generate a DSM of the full face and localized periorbital, perinasal, and perioral patches (fig. 2).

Following manual landmarking, collections of face surfaces were used to compute DSMs. A DSM is the set of modes resulting from a principal component analysis (PCA) of the residuals between a collection of "training" surfaces and their mean with the use of thousands of densely corresponded points. A more detailed description of the technique is provided elsewhere (Hutton et al. 2003; Hammond et al. 2004). Typically, ~100 PCA modes are sufficient to cover 99% of shape variation in

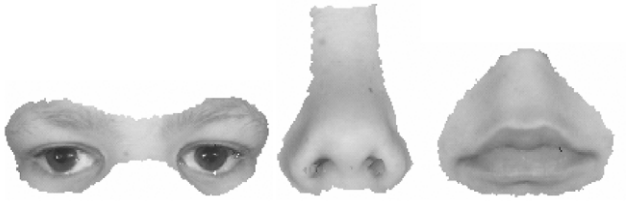

Figure 2 Illustration of the three localized face patches used in the discrimination testing. *Left,* periorbital; *center,* perinasal; and *right,* perioral.

a DSM of the full face. This subset of modes can be used to synthesize each face in the original training set. Thus, a surface originally represented by 20,000 points, or 60,000 variables, can be reconstructed using a weighted sum of ~100 modes. In the multidimensional "face-space" defined by the vectors of modes, the difference between two faces is computed as the Euclidean distance between the vectors representing them.

Delineation and Discrimination of Face Shape

For the discrimination testing, three pattern recognition algorithms—closest mean (CM), linear discriminant analysis (LDA), and support vector machines (SVM)—were used to classify the unseen test faces (Vapnik 1995). With CM, the average faces are computed for the control and syndrome subgroups in the training set, and each unseen test face is classified according to which average it is closer to. For LDA, the goal is a linear combination of PCA modes that exhibits the largest difference in the subgroup means relative to the within-group variance. SVM, or large margin classifiers, focus on individual cases in the overlap of the subgroups to be classified that help to define a separating surface with the largest margin between the subgroups.

The DSM representations of the faces in the 90% training set were presented to the three classification algorithms as a set of vectors to produce discrimination models for classifying the 10% test set unseen. Rather than selecting one of the associated sensitivity-specificity results arising from the 20 training-unseen test set pairs as representative (e.g., the one with the highest sensitivity), we constructed a receiver operating characteristics (ROC) curve for each split. The underlying distribution of position relative to the discriminating plane for CM discrimination has been verified as approximately normal for both the control and affected subgroups for the four syndromes considered. For each syndrome-control pair, and for the CM, SVM, and LDA pattern recognition algorithms, the 20 ROCs have been summarized by their mean. The ROC plots are available as additional material on the associated Eastman Dental Institute Web site.

For each ROC curve, we estimated the equal error rate, which is the point on the ROC curve where true positive and false positive rates sum to unity. For each algorithm, the classification performance of the 20 separate 90%–10% splits was summarized as the average of the equal error rates derived from the associated ROC curves. Henceforth, we refer to this as the "classification performance" and express it in terms of the corresponding true positive rate. For algorithms such as SVM, different model building parameter values and different kernel functions were used for training. The classification performance quoted in such instances is the best average for the different parameter combinations considered.

When evaluating the facial phenotype of particular individuals, we used the same models generated from the training sets to classify each individual's face, or face patch, unseen. This enabled an average classification position and 95% CIs to be estimated for each subject with respect to CM classification.

Dynamic morphs of faces and face patches varying between different control and syndromic group pairings provide an excellent delineation of shape differences. During the generation of the dense correspondence of points across a set of faces, it is necessary to trim back the surface to an area that is common to each surface in the set. Because the earlier face-scanning systems did not capture ears well, ears do not always feature in the DSMs generated. Although ear shape and position are important aspects in the gestalt recognition of faces, the classification performance based on face surfaces lacking ears are generally very accurate (typically >85%). Therefore, for DSMs computed solely for the visualization of facial form (see fig. 1), a subset of face scans with good ear coverage was selected and six additional landmarks were added to each face: preaurale and otobasion inferius for each ear and left and right frontotemporale. The latter two landmarks, even though invariably present in a captured face surface, were not used in the generation of DSMs for discrimination testing because of the unreliability inherent in their manual placement.

Clinical Summary of Four Adults with a Putative Diagnosis of WS

Four adult patients were selected as representing a broad spectrum of facial phenotypes in WS. Patient 1 did not appear to have obvious facial features of WS, whereas patient 2 had most of the features, relatively strongly. The face of patient 3 was thought to be typical, whereas patient 4 had some features but, overall, was atypical.

Patient 1 was referred as possibly having WS, mainly because of neonatal hypercalcaemia and developmental delay, but without strong WS facial morphology. He had no heart problems, and adult height was 176.1 cm. Patient 2 had problems in the neonatal period with failure to thrive. There was documented hypercalcaemia and constipation. He had no structural cardiac defect and motor milestones were delayed. Diagnosis was based on the typical personality and facial features. Patient 3 also had problems in the neonatal period with failure to thrive. A diagnosis was not suggested at this stage and there is no documented hypercalcaemia. There was no cardiac anomaly, but hyperacusis and learning delay were recorded. Diagnosis was based on personality and facial features typical for WS. Patient 4 had no docu-

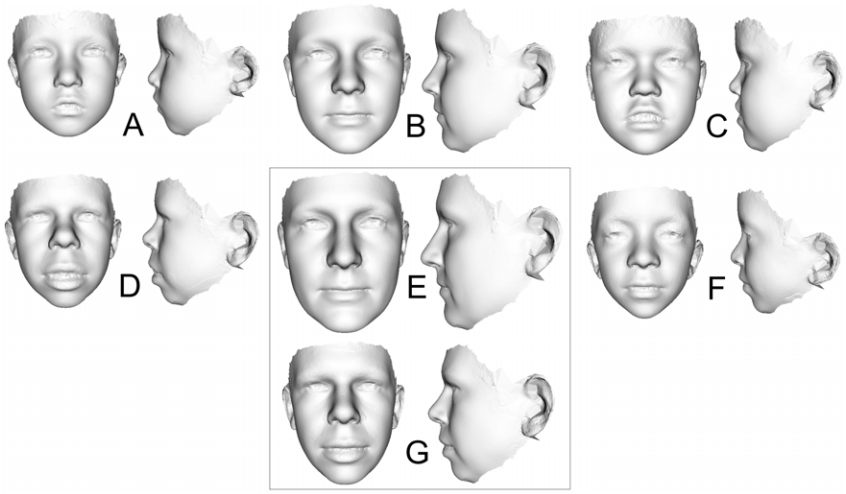

Figure 3 Portrait and profile views of average faces for the controls and for each syndrome (not to a common scale). *A*, 22q11DS; *B*, Control; *C*, SMS; *D*, WS; *E*, Control adults; *F*, NS; *G*, WS adults. Face shape differences are better evaluated by viewing the dynamic morphs on the Eastman Dental Institute Web site.

mented hypercalcaemia but was treated for bilateral inguinal herniae and developed hyperacusis later on. Developmental milestones were delayed. Photographs showed typical facial WS features in the early years, and a clinical diagnosis of WS was made at age 8.

Molecular Analysis

FISH was performed using a commercially available probe (ONCOR) to detect the 7q11.23 chromosome deletion. Microsatellite genotyping was performed to detect hemizygosity at 7q11.23 with the use of eight polymorphic markers (Généthon) mapping to the microdeleted region (Foster et al. 1993; Gyapay et al. 1994). PCRs were performed in mixtures containing 100 ng of genomic DNA, 750 μmol of each dNTP, 1μl 10 × PCR buffer at 37 mM Mg^{2+}, 5 pmol of each primer, and 0.25 units of *Taq* polymerase (BCL) in a final volume of 10 μl. Samples were processed through standard PCR conditions: denaturation at 94°C for 2 min, followed by 27 cycles of 93°C for 50 s, 58°C for 50 s (except *D7S2490* at 60°C), and 72°C for 1 min, with a final extension time of 5 min at 72°C. The PCR products were run on an 8% polyacrylamide gel (acrylamide: N,N′methylene bisacrylamide 19:1, 300V, 2 h 15 min), and the alleles were visualized by silver staining.

Results

Delineation of Average Face Shape Differences and Visualization of Average Growth

DSMs generated from a large set of children and young adults generally have a first principal component, or mode, that reflects overall face size and that correlates well with age, a typical Pearson product moment correlation coefficient being 0.9. Hence, the first mode of such a DSM is an approximation of facial growth. Separate DSMs for children's faces were computed for the control subset and for each of the four syndromes with the use of subsets of face scans for which there was good ear coverage. As described above, additional landmarks were added to these faces before the DSMs were computed. By morphing and simultaneously rotating the mean face, using a regression of age against mode 1 for each of these DSMs, we generated five sequences of facial growth (fig. 1).

Using the same subsets of images with good ear coverage and the extra landmarks, we computed four further DSMs for each of the control-syndrome comparisons. These models were used to produce static portrait and profile views of the average faces of the control subgroup and each group of individuals with a syndrome (fig. 3). Comparison of the static average faces provides further delineation of the facial differences. Besides these static growth sequences and averages, more revealing dynamic morphs can be viewed in the additional material provided online at the Eastman Dental Institute Web site.

Discrimination Testing

The results of the unseen discrimination testing using the three pattern recognition algorithms (CM, SVM, and LDA), are summarized in table 1. The discrimination performance of the full face for the 10 pairwise control-syndrome and syndrome-syndrome comparisons for children ranges from 87% to 98%, depending on the

Table 1

Summary of Discrimination Testing for Pairwise Comparisons of Controls and Each Syndrome

	Closest Mean (%)				Support Vector Machines (%)				Linear Discriminant Analysis (%)			
Comparison	Face	Eyes	Nose	Mouth	Face	Eyes	Nose	Mouth	Face	Eyes	Nose	Mouth
Children[a]:												
WS vs. Control	94	93	93	88	96	93	93	89	98	93	93	89
SMS vs. Control	87	89	94	88	88	93	94	89	91	92	94	86
NS vs. Control	94	89	81	83	94	92	82	86	94	89	82	83
22q11DS vs. Control	91	81	85	83	94	83	87	85	90	83	85	84
WS vs. NS	88	88	88	88	91	89	88	88	92	88	88	88
WS vs. SMS	88	91	90	88	94	95	95	88	94	96	94	88
WS vs. 22q11DS	89	90	89	89	91	94	96	91	90	95	94	90
NS vs. SMS	88	88	80	80	94	88	80	86	97	89	80	80
NS vs. 22q11DS	91	91	91	77	91	93	91	79	91	93	91	76
SMS vs. 22q11DS	91	91	91	82	91	93	92	85	90	91	91	85
Adults[b]:												
WS vs. Control	92	92	92	92	100	95	92	94	100	98	95	95

Note.—Each row of the table shows the discrimination rates for 1 of 10 pairwise comparisons of a control group and 4 syndrome populations. The first four columns show the classification performance for four face areas (full face, eyes patch, nose patch, and mouth patch) for the CM algorithm. The next four and the final four columns show the corresponding results for the SVM and LDC algorithms, respectively.

[a] Control ($n = 185$); WS ($n = 85$); NS ($n = 80$); SMS ($n = 54$); and 22q11DS ($n = 115$).

[b] Control ($n = 132$) and WS ($n = 45$).

algorithm, with LDA and SVM giving the best overall performance. Although the CM algorithm is relatively unsophisticated, its classification accuracy is close to that of LDA and SVM.

The control-syndrome comparisons in table 1 were also visualized in facial cartoons for more transparent interpretation (fig. 4). In each cartoon, the outside ellipse represents the full face and the inner ellipses represent the three face patches: periorbital, perinasal, and perioral. The darkness of the shading of each ellipse reflects the classification performance, with black representing the lowest value in the control-syndrome comparisons and white representing a classification performance of 100%. Thus, the darker the shading, the lower the classification performance. For example, the perfect discrimination between adult control faces and those of individuals with WS, an average classification performance of 100% (table 1), is reflected in the large ellipse being totally unshaded. The poorest classification performance, for the control-NS classification using the perinasal patch, is shown in black. Table 1 also summarizes the classification performance for each syndrome-syndrome comparison.

It is possible to illustrate face shape difference between controls and individuals with a syndrome by comparing an exaggeration of the mean face of a syndrome group with that of the mean control. In the discrimination analysis using CM, the relative positions of the mean control and syndrome group faces are normalized to the interval $[-1,1]$, and the position of a face along this interval, relative to the bisecting plane at its origin, determines its classification. The mean syndromic face, one unit distant from the origin, can be exaggerated to a position two units away. Typically, this is close to those individuals whose faces are the most dysmorphic. The surfaces of the faces of the syndrome mean exaggerated in this way and the mean control are then registered using an iterative closest-point procedure. The distance between the two registered surfaces is used to color code regions of the exaggerated face, with red showing those areas furthest from and inside the mean control face surface, and blue showing those areas furthest from and outside the mean control surface. Other colors, as in the accompanying scales, show regions of intermediate position. Figure 5 includes such "thermal" depictions of surface differences for NS (top left), SMS (top right), 22q11DS (bottom left), and WS (bottom right).

A 10-fold cross-validation has also been undertaken for full face discrimination testing with the CM classification of the combined syndrome subgroups. With the use of randomized 90%–10% training-test set pairs and maintaining stratification (proportional representation) across the four syndrome subgroups, the classification of a random collection of 320 affected children resulted in 35 misclassifications. Thus, the overall accuracy across all syndromes was 89%. The corresponding correct classification rates for each syndrome were as follows: NS (89%), SMS (78%), 22q11DS (89%), and WS (96%).

In the set of 185 controls used for each syndrome-

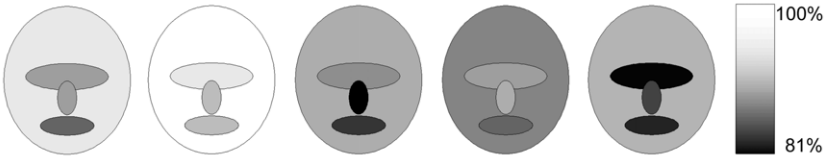

Figure 4 Alternative, cartoon representations of the control-syndrome discrimination performance in table 1. Each large ellipse represents the face, and the smaller inner ellipses represent the periorbital, perinasal, and perioral patches. The darkness of the shading of each ellipse reflects the classification performance, with black representing the lowest of the control-syndrome entries (at 81%), and white the highest. Thus, the darker the shading, the lower the classification performance.

based analysis of children's faces, 29 families were represented with ⩾2 siblings. No siblings were included in the adult data sets. To test sibling exclusion, we removed 37 siblings, leaving a control set of 148 unrelated children. We recomputed the discrimination performance for the unrelated controls and children with WS. The classification performance of the CM algorithm remained at 94%, whereas those of the LDA and SVM algorithms were reduced by 2%, from 98% to 96% and from 96% to 94%, respectively. We concluded that the removal of siblings makes little difference to the classification performance in WS. Similar calculations can be performed for the other entities described here.

3D Facial Analysis of Four Unconfirmed WS Patients

Figure 6 contains a background scatterplot of all 20 unseen, randomly generated test sets used to determine the classification performance of the nearest mean algorithm for adult controls and adults with WS. The vertical axis measures age in years. The horizontal axis measures the position of a face relative to the control and WS averages normalized to the interval [−1,1], with the average adult control face for the corresponding training set at −1 and the average face of the adults with WS at +1. The *Y*-axis can be thought of as the classification line. Thus, the scatterplot reflects the overall, and highly accurate, discriminating ability of the nearest mean classification algorithm for the full-face comparison of adult controls and adults with WS.

The average position of each selected adult's face relative to the control and WS mean faces over the 20 training sets was added to the background scatterplot. These average positions are also shown with 95% CIs. The same calculation of average positions relative to the two means was repeated for the 20 training sets used in the localized face patch analyses (fig. 7). Figure 7 shows the average positions and CIs for the same four adults for each face patch but without a background scatter of the unseen test set classification as in figure 6.

Molecular Genotypes of Four Adults with Putative WS Diagnosis

The molecular analysis is summarized in figure 8. Molecular diagnosis of WS by FISH testing identified a hemizygous deletion at the elastin locus in patients 2, 3, and 4. This was confirmed by microsatellite genotyping using polymorphic markers in the region. The classic WS deletion includes the markers *D7S489B–D7S1870*, but definition of the deletion size with the use of microsatellite mapping is limited by their polymorphic status in a given family and is generally used in diagnostic situations where FISH analysis is not possible or the results are unclear.

Patients 2 and 3 were deleted for the paternally inherited copy of the elastin gene and the microsatellite *D7S2472* (distal to the *CYLN2* gene), but uninformative for the flanking markers. Patient 4 was more informative, showing deletion of the maternally inherited copy of the elastin allele as well as the microsatellite markers in the genomic region spanning the genes *FZD9* to *GTF2I*. In contrast, patient 1 was not deleted for any of the informative microsatellite markers tested in the region (no sample was available for FISH analysis), suggesting that he does not have WS.

Comparison of 3D Face Analyses and Genotype of Adults with Putative WS Diagnosis

In patients 2 and 3, who fit the WS facial phenotype well, the molecular results identified a hemizygous deletion at 7q11.23. In the case of patient 4, the facial gestalt does not fit the classic WS phenotype. In the separate DSM analysis of his features, he showed positive for the full face, eyes, and nose, but negative for the mouth—the latter being consistent with a relatively short philtrum for WS. Molecular analysis confirmed that he did, indeed, have a microdeletion. Since patient 4 has been categorized as having less classic dysmorphism, he is undergoing more detailed analysis to determine if there are any differences at the molecular level. Finally, in patient 1, the positive result for the mouth is consistent with a relatively long philtrum. However, for the face,

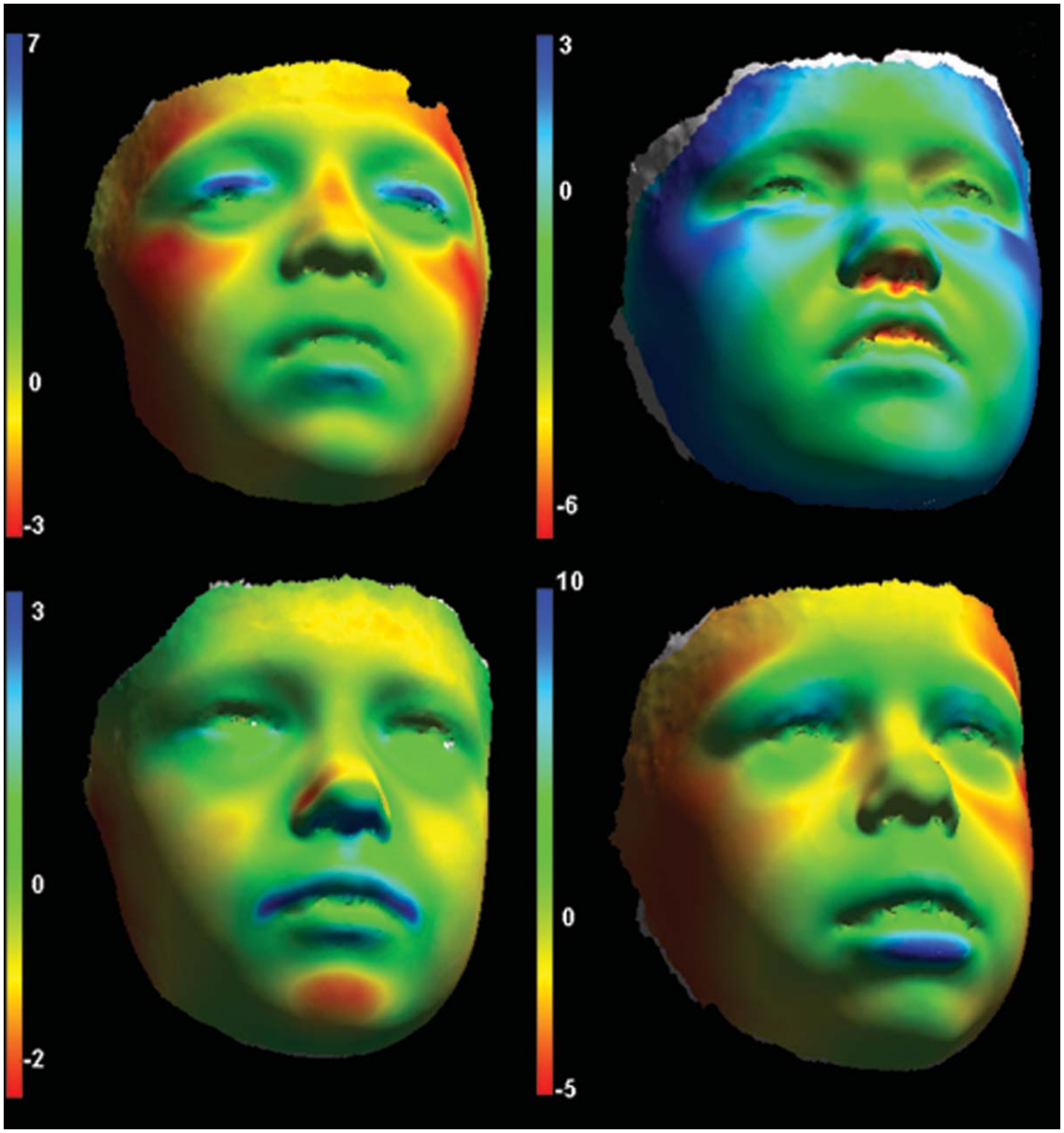

Figure 5 Comparison of exaggerated average faces of syndromic groups with the average control. *Upper left,* NS; *upper right,* SMS; *lower left,* 22q11DS; and *lower right,* WS. The color coding reflects the distance between the compared face surfaces. *Red,* regions most distant and internal to the mean control face. *Blue,* regions most distant and external. Other colors shown in the scales identify intermediate positions. Each figure has its own quantitative scale and also uses a different "zero" color for regions where the two surfaces coincide. It is not possible to employ a single scale to show both subtle and gross surface differences. Therefore, care must be taken in their interpretation.

eyes, and nose analysis, he was not categorized as having WS, which was confirmed by the lack of a gross genomic deletion in the critical region.

Discussion

Clinical geneticists often refer to facial differences and features that offer clues to diagnosis. Articulating those differences can be challenging, especially if an individual's face shows a mild facial phenotype or if morphological characteristics are manifest in only part of the face. Our results suggest that static and dynamic visualizations derived from DSMs of 3D images can delineate dysmorphic features well and, hence, may be useful in the training of clinical geneticists and pediatricians.

Being able to test the discriminating ability of specific regions is useful for anatomical regions of particular interest to clinicians or for tissues identified as being

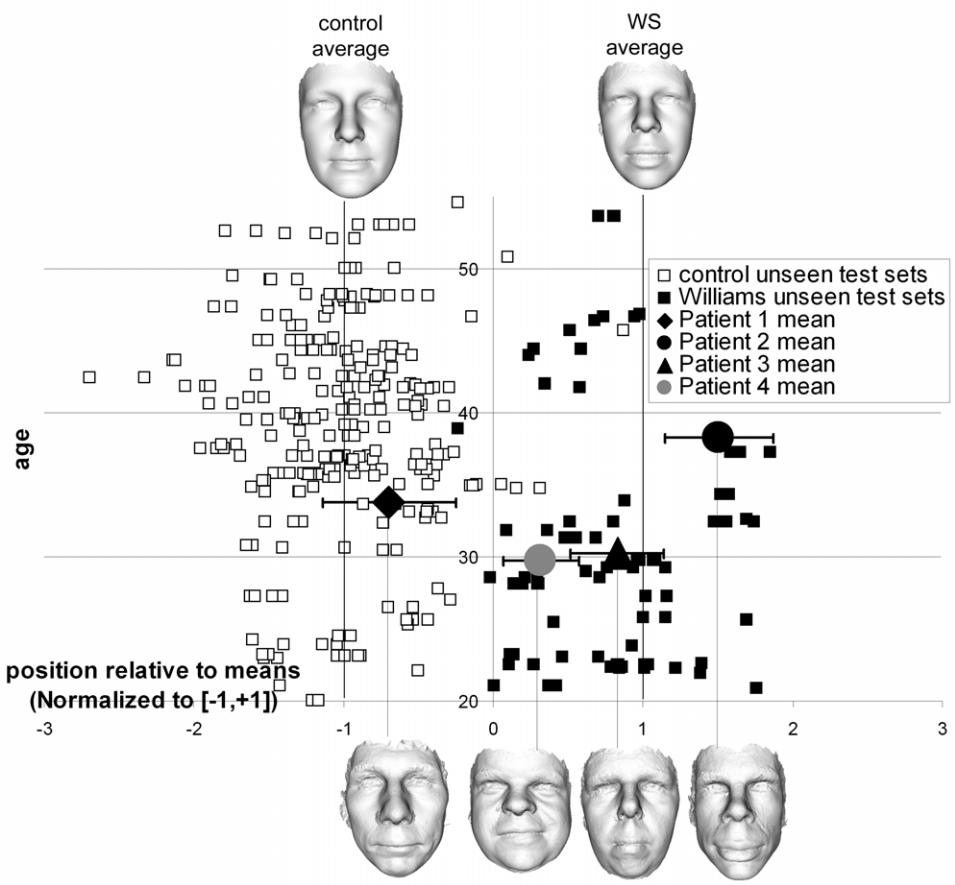

Figure 6 Average classification position between control and WS mean faces, for faces of four selected adults with WS for 20 DSMs computed from the randomized 90%–10% splits used to calculate average discrimination performance. CIs of 95% have been added to the average classification position. The background scatterplot shows the classification position of the 20 unseen test sets for the relevant DSM. The vertical broken lines mark the position of the two means at −1 and +1 respectively. All other positions are normalized to [−1,+1].

developmentally important, perhaps in related mouse studies. In-depth analysis to identify the discriminating features associated with each syndrome has highlighted differences in the discrimination performance of localized patches of the face (periorbital, perinasal, and perioral). It can be seen from the face discrimination cartoons (fig. 4) or from table 1 that for the WS-control comparison for children, the nose and eyes patches discriminate very well, almost as well as the full face. The mouth patch discriminates less well. By comparison, the adult results for the mouth are much better, which is not surprising, since this is a feature that is exaggerated with age. The documented increase in typical features of WS in later life underlies the extremely high discrimination rates for adults, especially for the full face at or close to 100%.

The NS-control discrimination was noticeably poorer for the mouth and nose patches and was not much better for the eyes patch. The photogrammetric scanners used in this study are known to have problems capturing wet surfaces and hair. Therefore, less accurate capture of eye surfaces, eyelashes, and eyebrows may diminish facial differences, such as downward sloping palpebral fissures and ptosis of the eyelids, that are obvious to the naked eye. We conclude that consideration of the entire face, the overall gestalt, is much more important in NS than in, say, WS.

The most intriguing of the discrimination results are those for SMS. The eyes and nose patches perform reasonably well, considering the comparatively small number of affected individuals included. The unusual shape of the upper lip frequently features in the description of the gestalt for SMS and can be easily observed in the static comparison of average faces (fig. 3) or, more dramatically, in the dynamic morph provided in the additional material. However, the eyes and nose patches discriminate at a significantly higher level. For the eyes patch, the midfacial hypoplasia and flatness of the nasal bridge may be the key. The nose patch overlaps at the nasal bridge but also reflects the significant upward and backward deflection of the subnasale, often resulting in a hidden columella and concave undersurface to the

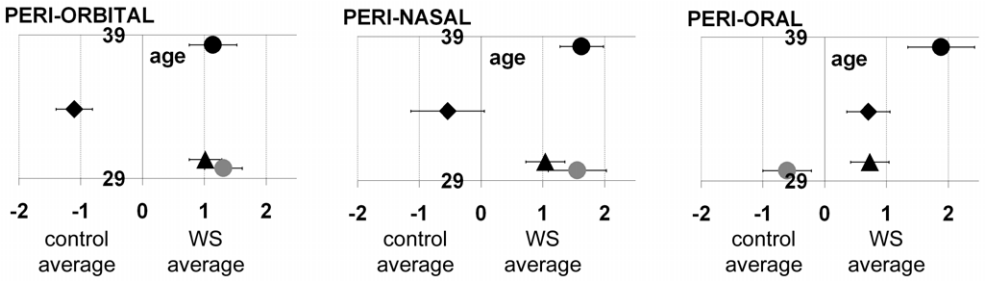

Figure 7 Average classification of four selected individuals with WS for the three localized face patches

nose. The poorer performance of the full face is similar to that for the mouth patch. One possible explanation may be due to the nose and eyes already discriminating well. So, the increase in the area examined to the full face, and the concomitant increase in the number of variables analyzed, occur without an increase in the number of training examples to generate a better discriminating model. However, such an explanation may not sit well with the WS discrimination performance, where the eyes and nose patches also do well.

Given the relatively subtle facial phenotype in 22q11DS, the better performance of the full face over the nose and eyes was expected. The narrow mouth and shallow groove between the lower lip and chin (the latter not usually included in the gestalt description) may explain the reasonable perioral performance. However, lip tension and facial expression in subjects with 22q11DS cannot be discounted as an influencing factor. Even so, the full face discriminates at a satisfactory level.

Since our first 3D face-based analysis of NS and 22q11DS, more accurate laser and photogrammetric scanners have become available and more families have been recruited. As a result, the previous visualizations of face shape differences (Hammond et al. 2003*b*) have been dramatically improved, as have the discrimination results for the full face. Because the recruitment of affected individuals continues to be relatively low, male and female patients have had to be combined in the delineation modeling and discrimination testing. However, as recruitment improves, separation of the populations by both gender and finer chronological steps will be possible and should further improve the results.

The potential of using DSMs of face shape to aid clinical diagnosis was also highlighted in this study. The syndrome-syndrome discrimination results are sufficiently accurate to encourage the use of DSM-based techniques for screening individuals simultaneously for a range of facial phenotypes. The molecular analysis of the four patients, three with clinical descriptions suggestive of WS, confirms the accuracy of the predictions made by the WS adult models and supports their use in a diagnostic environment. Moreover, such accuracy in pinpointing similarities and differences in facial morphology suggests that DSMs may have a future role in the investigation of patient subgroups who share a phenotypic or genotypic description. 3D dense surface modeling allows objective grouping of affected individuals into those who fit the classic facial phenotype, those who clearly do not, and those who justify more detailed molecular mapping, such as microarray comparative genome hybridization, to investigate genotype-phenotype correlations.

In the discrimination testing, particular localized regions of the face were used. From the discrimination analysis of the whole face surface, it is also possible to derive and visualize by color coding the parts of the face that differ the most. Careful examination of the "thermal" images shown in figure 5 provides insight into the origins of face shape difference. For example, the figure panel for WS (bottom right) shows that the prominence around the mouth is partly a real prominence in the premaxilla, maxilla, and mandible. In part, however, it also arises from hypoplasia of the malar region. Similarly, the prominent supraorbital arches are, in part, caused by true hyperplasia of that area, but also, in part, by hypoplasia of the superior parts of the forehead.

In conclusion, dense surface modeling of the full face and of localized facial morphology supports clinical training and the screening of diagnostic options and can contribute to phenotype-genotype analysis. The diagnosis of WS may not be a challenge to experienced clinicians, but objective computer-based face shape analysis is likely to be particularly helpful in evaluating partial deletion patients and understanding facial change over developmental time, in situations where there is a lack of local clinical expertise or a lack of funds and/or facilities for genetic testing or where no appropriate genetic test exists.

Acknowledgments

The authors are very grateful to the individuals and families who consented to have 3D photographs taken. The following organizations were extremely helpful in facilitating the re-

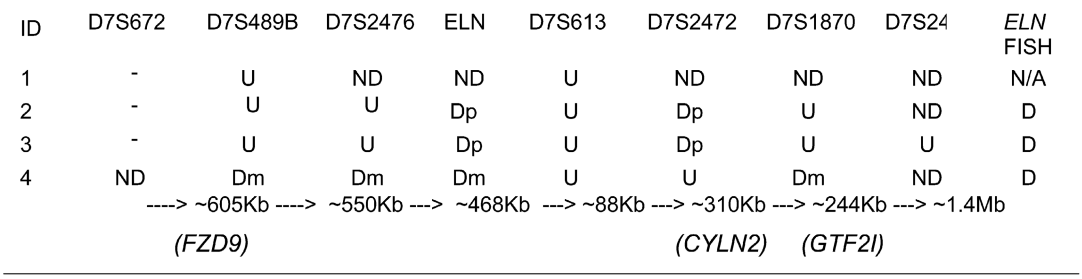

Figure 8 Summary of microsatellite analysis of polymorphic markers in the chromosome 7q11.23 deleted region in four putative WS patients. Note that genes around the microsatellite loci as well as the approximate distance between markers are shown at the bottom of the figure. ID = Patient identifier. U = Uninformative. ND = Not deleted. N/A = Not available. D = Deleted. p = Paternal allele. m = Maternal allele.

cruitment of families with affected children: for NS, TNNSG and NewLife/Birth Defects Foundation; for 22q11DS, Great Ormond Street Hospital Cleft Palate Clinic, MAXAPPEAL, The 22q11 Group (UK), and VCFSEF (USA); for SMS, SMS Foundation (UK) and PRISMS (USA); for WS, WSA (USA) and WSF (UK). We also acknowledge the recruitment of controls through volunteers at collaborating institutions and the St. Albans Clinic in London. NewLife (Birth Defects Foundation, grant number 2000/27) provided funding for two photogrammetric scanners and some travel expenses. M.T. is funded by Wellcome Trust grant number 061183. Wellcome Trust also provided a travel grant to P.H. for attendance at a WSA National Conference. P.H., A.K.-S., and M.T. received funding from the National Institutes of Health (NIH) (Fogarty/NIH grant R21TW06761-01). The NIH National Institute of Dental and Craniofacial Research Dental Clinic provided images of subjects enrolled in the SMS natural history study (NIH protocol 01-HG-0109). Statistical advice was provided by Dr. Aviva Petrie, and the article has benefitted significantly from the detailed comments of an anonymous reviewer.

Web Resources

Accession numbers and URLs for data presented herein are as follows:

Eastman Dental Institute, http://www.eastman.ucl.ac.uk/~dmi/faceparts.htm/ (for dynamic morphs of mean faces)
Eastman Dental Institute ROC plots, http://www.eastman.ucl.ac.uk/~dmi/roc_plots.htm/ (for discrimination analysis)
Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nih.gov/Omim/ (for NS, 22q11DS, Bardet-Biedl syndrome, SMS, and WS)

References

Allanson JE, Cole TRP (1996) Sotos syndrome: Evolution of the facial phenotype subjective and objective assessment. Am J Med Genet 65:13–20

Allanson JE, Greenberg F, Smith ACM (1999) The face of Smith-Magenis syndrome: a subjective and objective study. J Med Genet 36:394–397

Allanson JE, Hall JG, Hughes HE, Preus M, Witt RD (1985) Noonan syndrome: the changing phenotype. Am J Med Genet 21:507–514

Allanson JE, Hennekam RCM (1997) Rubinstein-Taybi syndrome: objective evaluation of craniofacial structure. Am J Med Genet 71:414–419

Allanson JE, O'Hara, Farkas LG, Nair RC (1993) Anthropometric craniofacial pattern profiles in Down syndrome. Am J Med Genet 47:748–752

Beales PL, Warner AM, Hitman GA, Thakker R, Flinter FA (1997) Bardet-Biedl syndrome: a molecular and phenotypic study of 18 families. J Med Genet 34:92–98

Foster K, Ferrell R, King-Underwood L, Povey S, Attwood J, Rennick R, Humphries SE, Henney AM (1993) Description of a dinucleotide repeat polymorphism in the human elastin gene and its use to confirm assignment of the gene to chromosome 7. Ann Hum Genet 57:87–96

Gorlin RJ, Cohen MM, Hennekam RCM (2001) Syndromes of the head and neck. Oxford University Press, New York

Greenberg F, Smith ACM, Richter S, Magenis E, Guzzetta V, Patel PI, Lupski JR (1991) Molecular analysis of the Smith-Magenis syndrome: a possible contiguous-gene syndrome associated with del(17)(p11.2). Am J Hum Genet 49:1207–1218

Gyapay G, Morissette J, Vignal A, Dib C, Fizames C, Millasseau P, Marc S, Bernardi G, Lathrop M, Weissenbach J (1994) The 1993–94 Généthon human genetic linkage map. Nat Genet 7:246–339

Hammond P, Hindocha N, Hutton TJ, Beales PL (2003*a*) 3D dense surface modeling defines a characteristic facial phenotype in Bardet-Biedl syndrome. Am J Hum Genet 73(S1): 284

Hammond P, Hutton TJ, Allanson JE, Campbell LE, Hennekam RCM, Holden S. Murphy KC, Patton MA, Shaw A, Temple IK, Trotter M, Winter RM (2004) 3D analysis of facial morphology. Am J Med Genet A 126:339–348

Hammond P, Hutton TJ, Allanson JA, Smith ACM (2003*b*) The 3D face of Smith-Magenis syndrome (SMS): a study using dense surface models. Eur J Hum Genet 11(S1):102

Hutton TJ, Buxton BF, Hammond P, Potts HWW (2003) Estimating average growth trajectories in shape-space using kernel smoothing. IEEE Trans Med Imaging 22:747–753

Jerome LA, Papaioannou VE (2001) DiGeorge syndrome phenotype in mice mutant for the T-box gene, Tbx1. Nat Genet 27:286–291

Loos HS, Wieczorek D, Würtz RP, von der Malsburg C, Horthemke B (2003) Computer-based recognition of dysmorphic faces. Eur J Hum Genet 11:555–560

Morris CA, Demsey SA, Leonard CO, Dilts C, Blackburn BL (1988) Natural history of Williams syndrome: physical characteristics. J Pediatr 113:318–326

Scambler PJ, Kelly D, Lindsay E, Williamson R, Goldberg R, Shprintzen R, Wilson DI, Goodship JA, Cross IE, Burn J (1992) Velo-cardio-facial syndrome associated with chromosome 22 deletions encompassing the DiGeorge locus. Lancet 339:1138–1139

Sharland M, Morgan M, Patton MA (1993) Photoanthropometric study of facial growth in Noonan syndrome. Am J Med Genet 45:430–436

Slager RE, Newton TL, Vlangos CN, Finucane B, Elsea SH (2003) Mutations in RAI1 associated with Smith-Magenis syndrome. Nat Genet 33:466–468

Tartaglia M, Mehler EL, Goldberg R, Zampino G, Brunner HG, Kremer H, van der Burgt I, Crosby AH, Ion A, Jeffery S, Kalidas K, Patton MA, Kucherlapati RS, Gelb BD (2001) Mutations in *PTPN11*, encoding the protein tyrosine phosphatase SHP-2, cause Noonan syndrome. Nat Genet 29: 465–468

Vapnik V (1995) The nature of statistical learning theory. Springer, New York

Yagi H, Furutani Y, Hamada H, Sasaki T, Asakawa S, Minoshima S, Ichida F, Joo K, Kimura M, Imamura S, Kamatani N, Momma K, Takao A, Nakazawa M, Shimizu N, Matsuoka R (2003) Role of TBX1 in human del22q11.2 syndrome. Lancet 362:1366–1373

Ward RE, Jamison PL, Allanson JE (2000) Quantitative approach to identifying abnormal variation in the human face exemplified by study of 278 individuals with five craniofacial syndromes. Am J Med Genet 91:8–17

Winter RM (1996) What's in a face? Nat Genet 12:124–129