

AI Facial Phenotyping: From Pattern Recognition to Clinical Evidence

Dr. Tzung-Chien Hsieh

Institute for Genomic Statistics and Bioinformatics

University Hospital Bonn

Manchester ESHG Syndromology & Dysmorphology Course

16th October 2026



nature

An agentic system for rare disease diagnosis with traceable reasoning

Zhao et al. 2026

Rare-disease diagnosis

nature

Towards conversational diagnostic artificial intelligence

Tu et al. 2025



Advancing conversational diagnostic AI with multimodal reasoning

Saab et al. 2026

nature

Towards conversational artificial intelligence for disease management

Liévin et al. 2026

nature

Towards accurate differential diagnosis with large language models

Mcduff et al. 2025

From Clinical Gestalt to Quantitative Evidence



Recognizing facial patterns

From clinical expertise to computational facial analysis



Quantifying facial gestalt

Cohort-level similarity and syndrome delineation



Translating similarity into evidence

Patient-level analysis and ACMG/AMP PP4 calibration

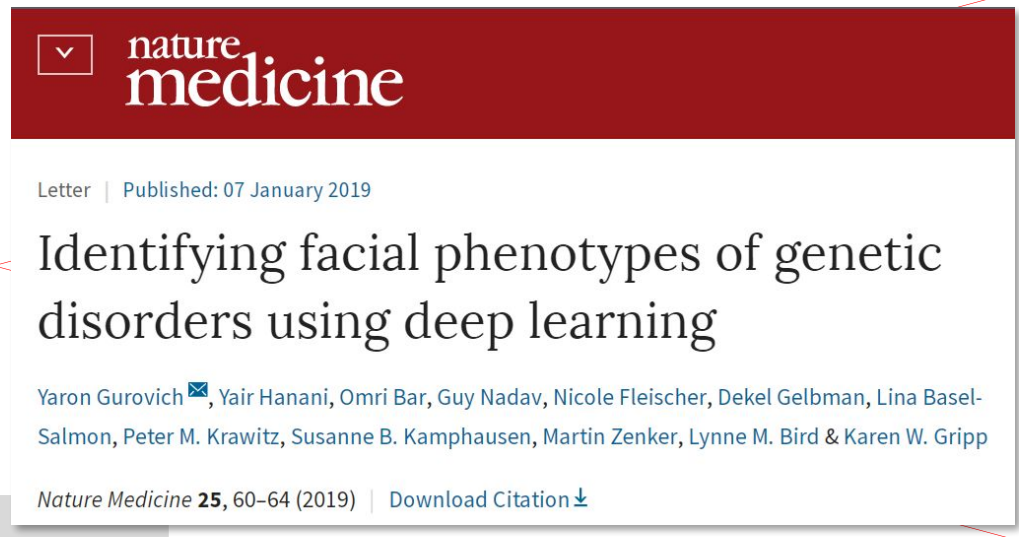


Bringing AI into clinical practice

Ancestry, generalization, and clinical implementation

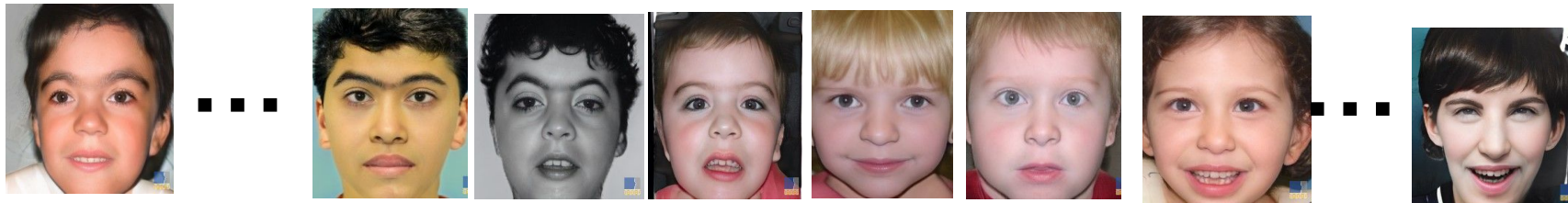
DeepGestalt Approach (Gurovich *et al. Nat Med* 2019)

DeepGestalt



FACE2GENE

↑ Training



GestaltMatcher Approach (Hsieh and Bar-Haim *et al.* *Nat Genet* 2022)

DeepGestalt

GestaltMatcher



Technical Report | Published: 10 February 2022

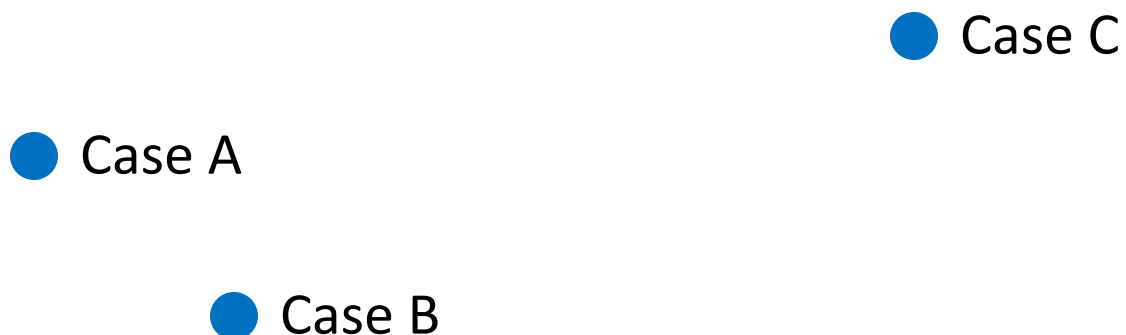
GestaltMatcher facilitates rare disease matching using facial phenotype descriptors

Tzung-Chien Hsieh, Aviram Bar-Haim, ... Peter M. Krawitz 

[+ Show authors](#)

What a facial score measures

Illustrative case relationships



A and B have similar
model representations

Patient similarity

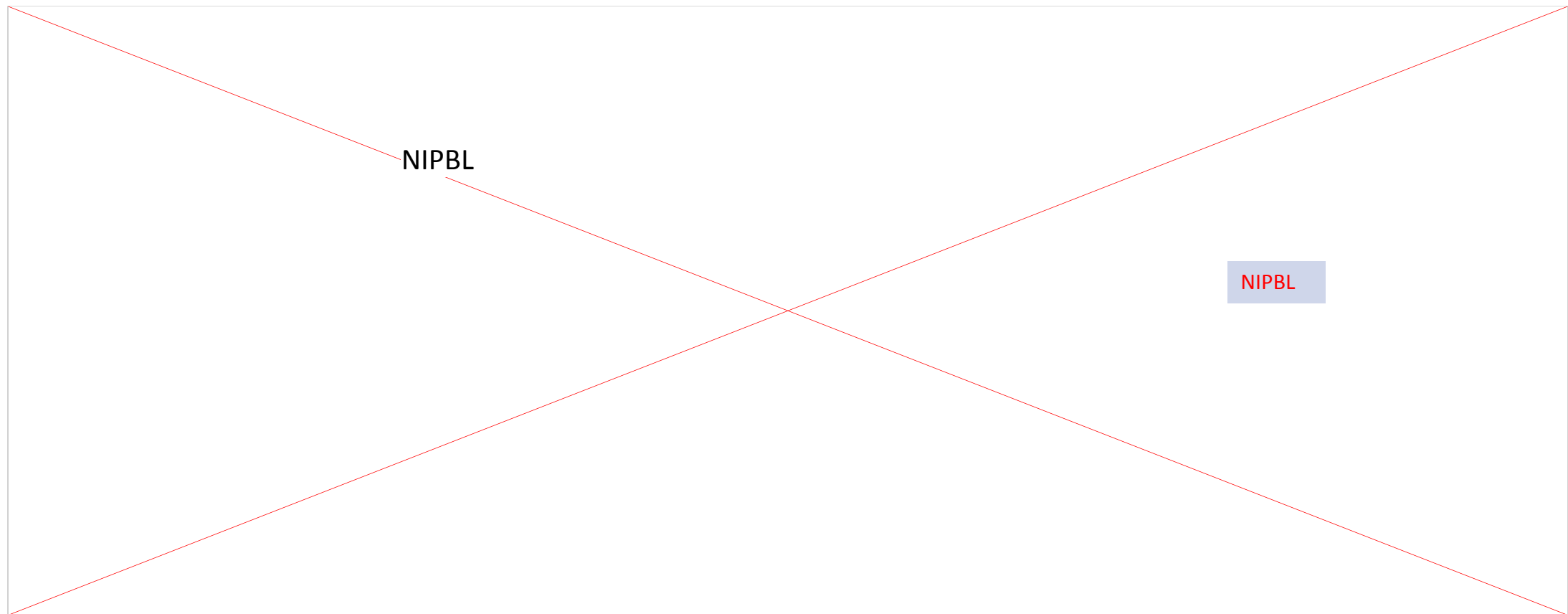
Model-dependent
resemblance
to reference cases

Ranked differential

Calibration is a further step
before phenotype evidence

Schematic only. Similarity is computed in model space, not from these positions.

Integrating GestaltMatcher scores into the exome analysis



PP4: phenotype specificity

A phenotype or family history highly specific for a disease with a single genetic etiology

Facial resemblance informs the phenotype assessment
Variant interpretation also requires molecular evidence



Cohort similarity and subgroup delineation

PSMC5

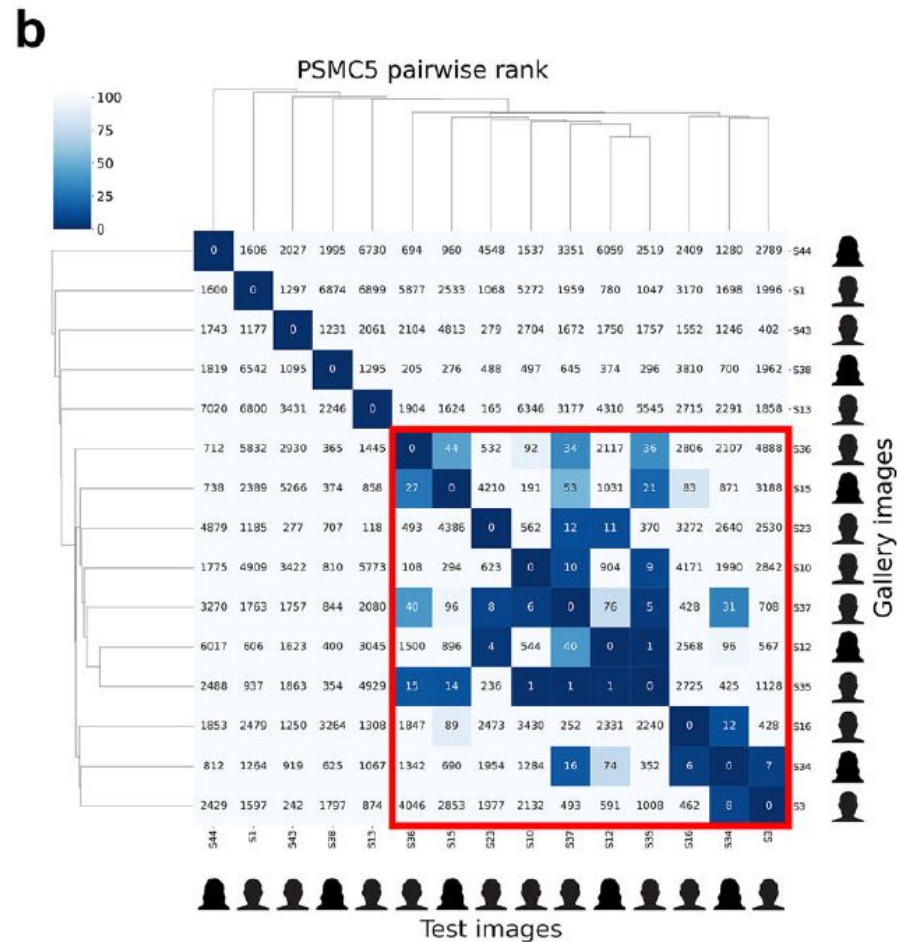
Do these patients share
a measurable facial gestalt?

MN1

Do variant-associated groups
show different facial
phenotypes?

Within-cohort resemblance and between-group separation need different comparisons

PSMC5: a heterogeneous clinical cohort

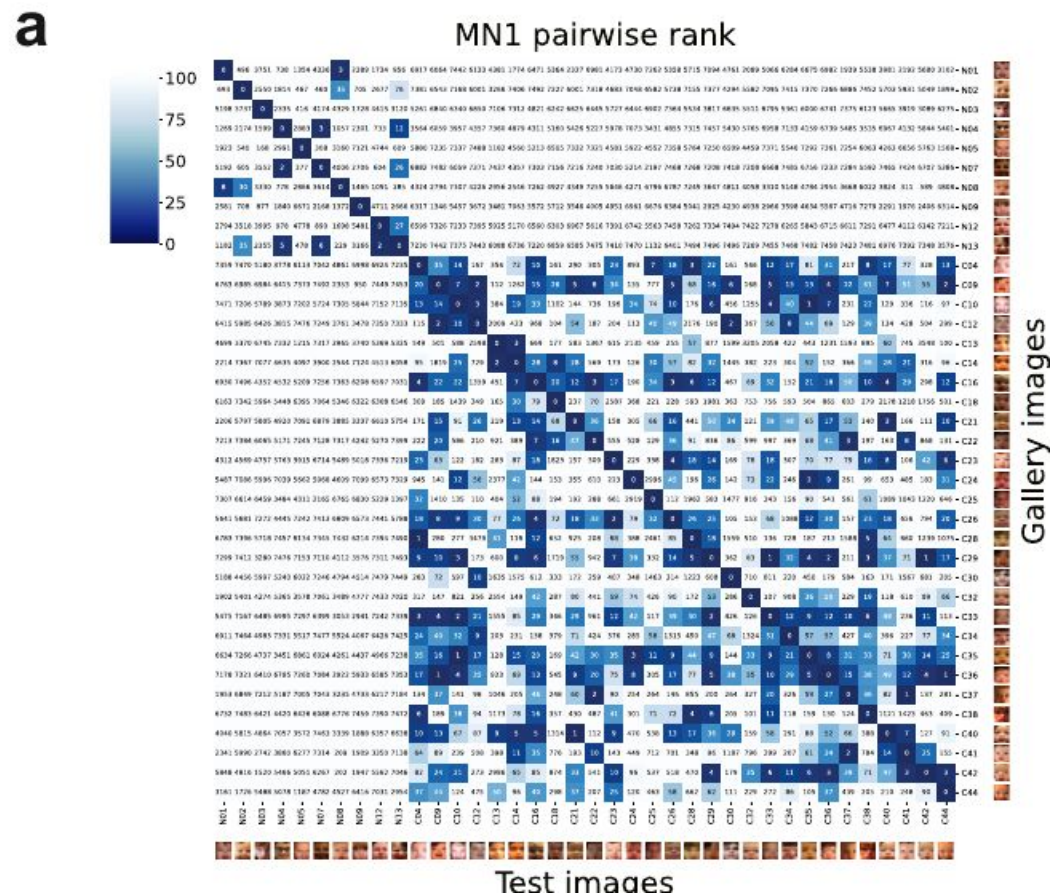


44 clinical participants
15 in facial analysis

Clinical appearance is
heterogeneous

Pairwise ranks reveal shared
resemblance

MN1: two phenotypes within one gene



56 individuals in the cohort

38 facial images

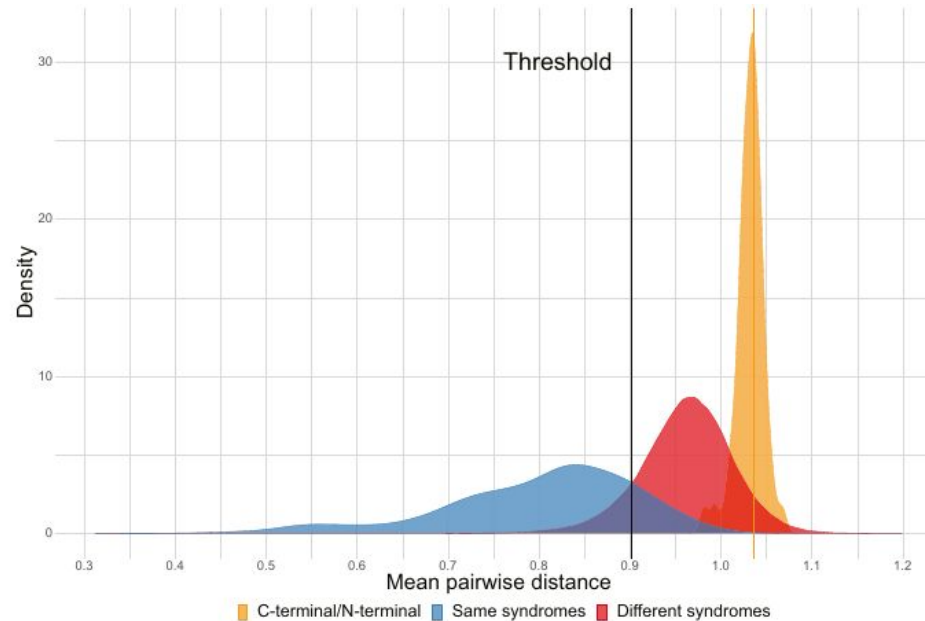
28 C-terminal truncations

10 N-terminal truncations

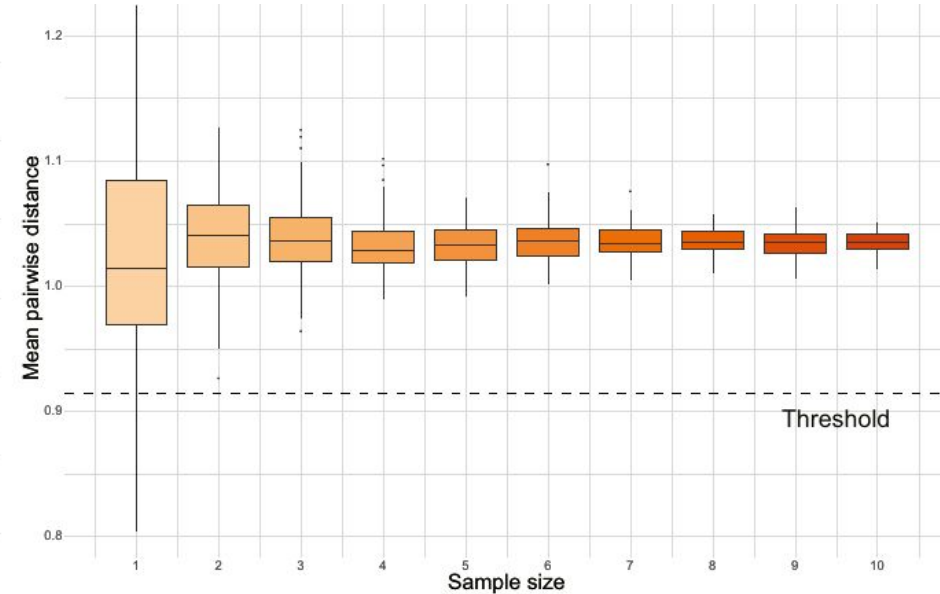
Variant-associated groups show distinct facial patterns

MN1: evidence for distinct subgroups

c



d

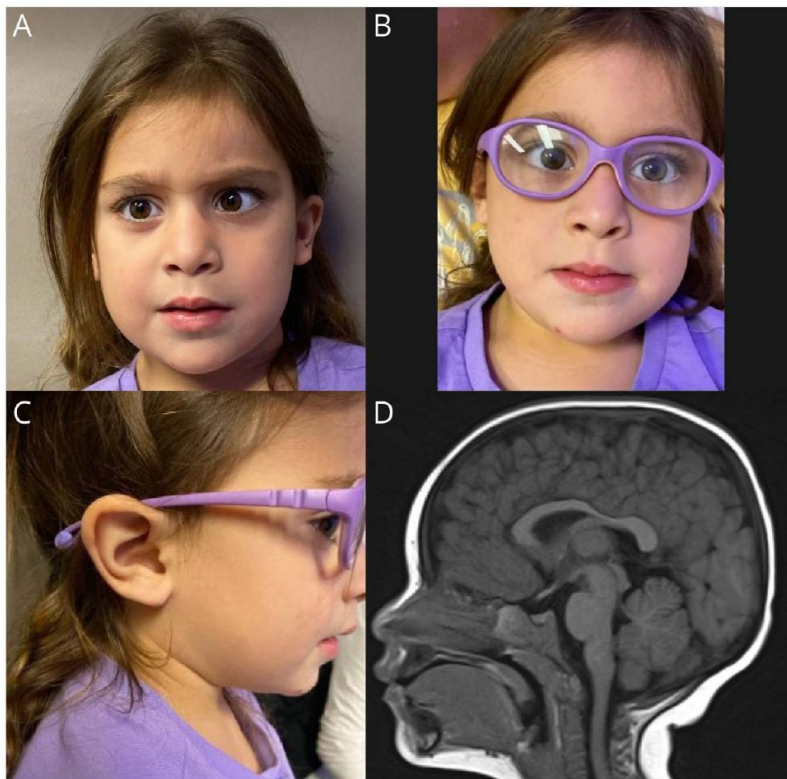


Between-group distance 1.035
Study threshold 0.901

Facial separation complements
independent DNAm evidence

Mowat–Wilson syndrome: the clinical case

Figure 2 Facial Profile, Feet, and T1 Sagittal Slice of the MRI Brain of the Proband



(A and B) Facial photographs of the proband. (C) The profile photograph of the proband. The facial and profile photographs show esotropia, short columella, broad nasal tip, low-set ears, posteriorly rotated ears, uplifted earlobe, broad forehead, and thick eyebrows. (D) Midsagittal MRI of the brain showing hypoplasia of the body of the corpus callosum.

4-year-old proband ZEB2 last-exon frameshift

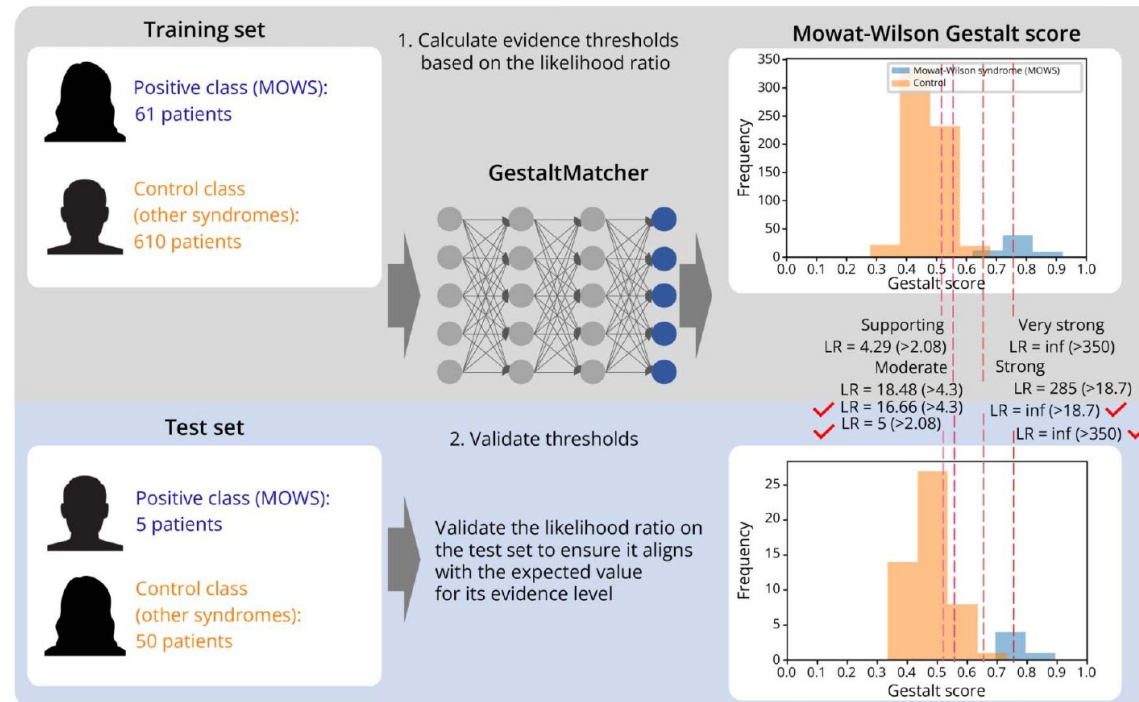
Initial laboratory VUS

Clinical suspicion of MWS

MWS ranked first by facial
analysis

Mowat–Wilson syndrome: score calibration

Figure 1 Derivation and Validation of Gestalt Score Thresholds for Mowat-Wilson Syndrome (MOWS)



Workflow for mapping the GestaltMatcher facial gestalt score to likelihood ratios (LRs) that meet ACMG/AMP evidence strength bins and validating the thresholds on a held-out set. (Top, training) Positive class: MOWS, $n = 61$; control class: other syndromes, $n = 610$. Histograms show the distribution of Gestalt scores for cases (blue) and controls (orange). Vertical dashed lines denote score thresholds corresponding to ACMG evidence levels: supporting (target $LR \geq 2.08$), moderate (≥ 4.3), strong (≥ 18.7), and very strong (≥ 350). On the training data, the observed LRs at these thresholds met or exceeded the targets (e.g., supporting $LR \approx 4.29$; moderate $LR \approx 18.48$; strong and very strong $LR \rightarrow \infty$ because no controls exceeded the thresholds). (Bottom, test) Positive class: MOWS, $n = 5$; controls: other syndromes, $n = 50$. Applying the same thresholds yielded LRs consistent with their intended evidence levels (e.g., supporting $LR \approx 5.0$ and moderate $LR \approx 16.7$), confirming that these thresholds also work in the test set.

Syndrome-specific comparison data

61 MWS / 610 controls
5 MWS / 50 controls
in the stated holdout

Small calibration sets
limit evidence strength

Four photographs and phenotype evidence

Photograph	Gestalt score	Study PP4 level
1	0.593	Moderate
2	0.580	Moderate
3	0.647	Moderate
4	0.547	Supporting

Supporting ≥ 0.5111 Moderate ≥ 0.5672

Repeated photos of one patient contribute correlated phenotype information

Challenges

THE EXPECTATION:

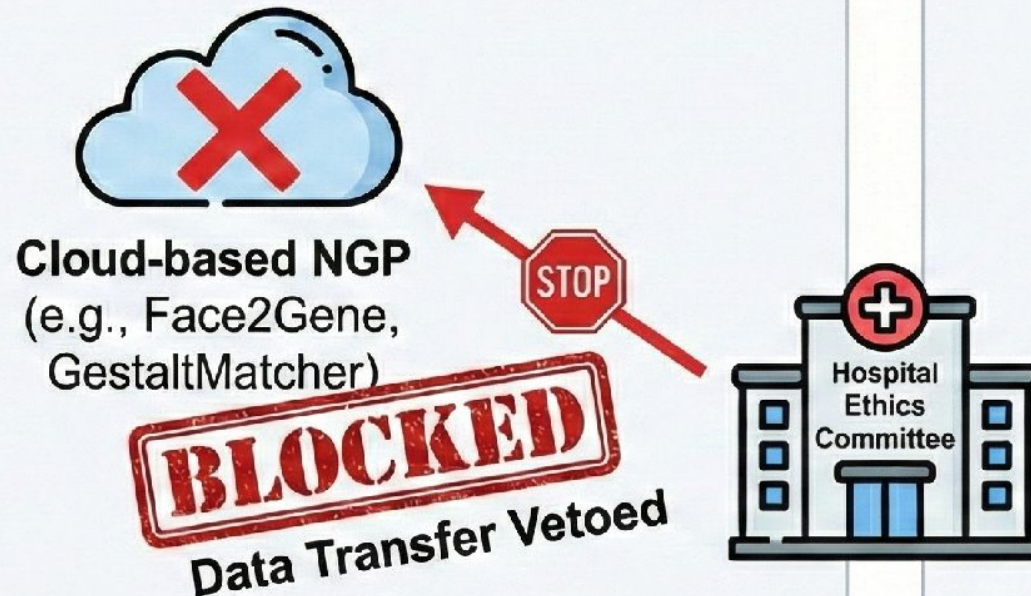


**THE REALITY
(ESPECIALLY IN ASIA):**



The privacy barrier in practice

The Cloud Dilemma



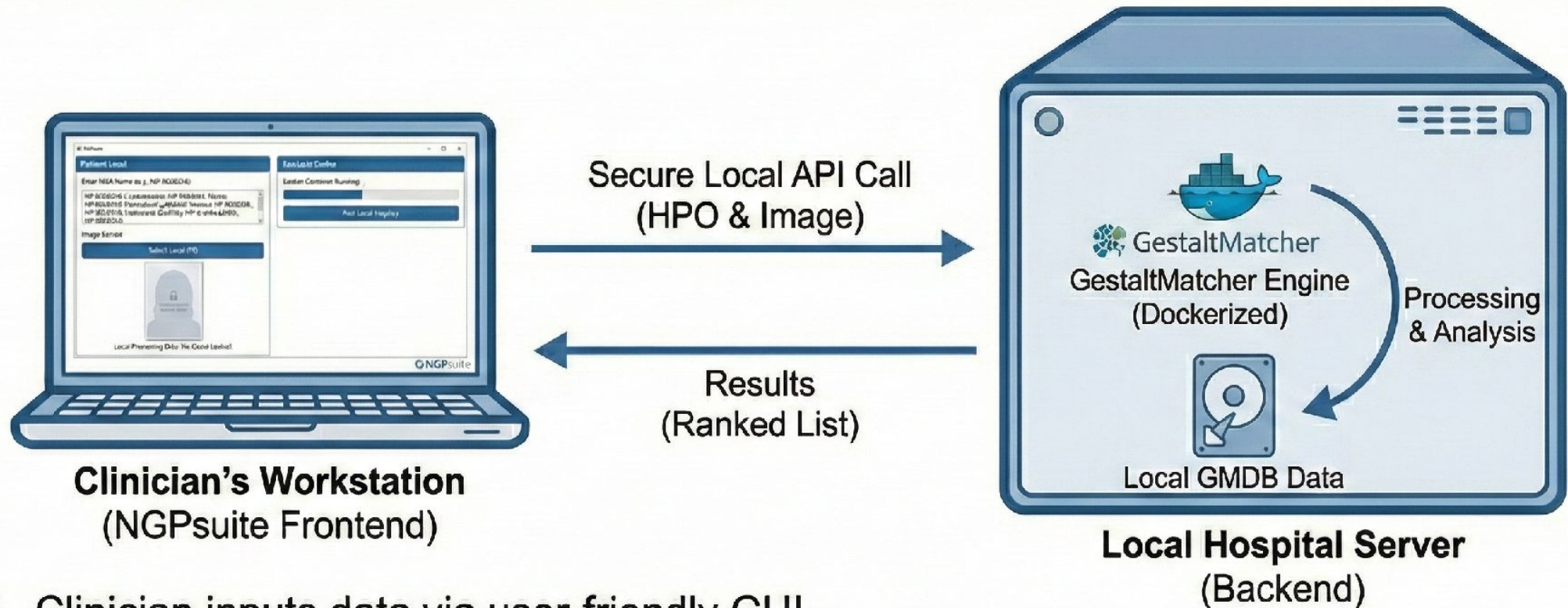
Mishima *et al.* (Japan): Ethics committees often veto external transfer of identifiable facial images.

The Local Solution (Motivation)



Motivates need for local, secure solutions that keep data within the institution.

NGPsuite bridges the clinician and AI



- 1. Clinician inputs data via user-friendly GUI.
- 2. NGPsuite sends secure request to local backend over hospital network.
- 3. Dockerized GestaltMatcher processes data entirely on-site.
- 4. Ranked results are returned to the GUI for review.

Running NGPsuite on your computer

C

Queried HPO

HPO ID	Name
HP:0008897	Postnatal growth retardation
HP:0001249	Intellectual disability
HP:0001007	Hirsutism

Meta Rank	GM Rank	PCF Rank	Syndrome Name	OMIM ID	GM Distance	GM Match	PCF Score	PCF Match
1	1	1	Cornelia de Lange syndrome	122470 ☐	0.4760		0.9780	
2	6	2	Coffin-Siris syndrome	135900 ☐	0.6730		0.9583	
3	3	23	OGDEN SYNDROME; OGDNS	300855 ☐	0.6190		0.8622	

1	1	1	Cornelia de Lange syndrome	☐	0.4760		0.9780	
2	6	2	Coffin-Siris syndrome	135900 ☐	0.6730		0.9583	
3	3	23	OGDEN SYNDROME; OGDNS	300855 ☐	0.6190		0.8622	
4	22	4	Rubinstein-Taybi syndrome	180849 ☐	0.7300		0.9042	
5	10	37	INTELLECTUAL DEVELOPMENTAL DISORDER, AUTOSOMAL DOMINANT	615761 ☐	0.6970		0.8450	

Output of NGPsuite

Meta Rank = Average of GestaltMatcher and PubCaseFinder Rank



GestaltMatcher Rank



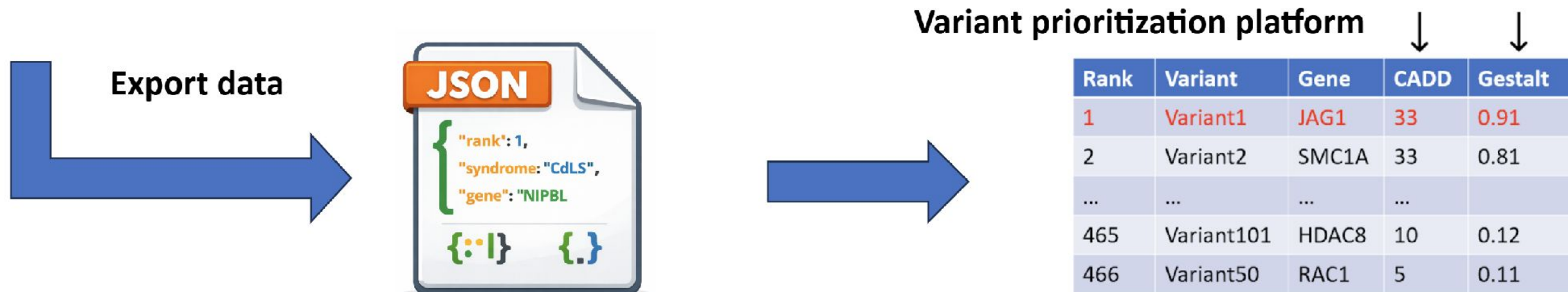
PubCaseFinder Rank



Meta Rank	GM Rank	PCF Rank	Syndrome Name	OMIM ID	GM Distance	GM Match	PCF Score	PCF Match
1	1	1	Cornelia de Lange syndrome	122470 	0.4760		0.9780	
2	6	2	Coffin-Siris syndrome	135900 	0.6730		0.9583	
3	3	23	OGDEN SYNDROME; OGDNS	300855 	0.6190		0.8622	

Integrate with your variant prioritization pipeline

Meta Rank	GM Rank	PCF Rank	Syndrome Name	OMIM ID	GM Distance	GM Match	PCF Score	PCF Match
1	1	1	Cornelia de Lange syndrome	122470 ☑	0.4760		0.9780	
2	6	2	Coffin-Siris syndrome	135900 ☑	0.6730		0.9583	
3	3	23	OGDEN SYNDROME; OGDNS	300855 ☑	0.6190		0.8622	



GestaltMatcher Database

[Home](#) [Gallery](#) [Groups](#) [My Data](#) [Downloads](#) [Publications](#) [About](#)

tzung.chien.hsieh@gmail.com



Search Gene name, OMIM ID, or PubMed ID

All



Search



Access

<https://db.gestaltmatcher.org/>

[Home](#) [Downloads](#) [Publications](#) [About Us](#) [FAQs](#)

[Sign up](#) [Login](#)



GestaltMatcher Database

A collection of curated medical photography of genetic syndromes as a resource for clinician and computer scientists.

[Log In](#)

To sign up, please send an email to
info@gestaltmatcher.org



Gallery View

View all in the database in the gallery view at a glance - you can search by gene, phenotypic series or disorder.

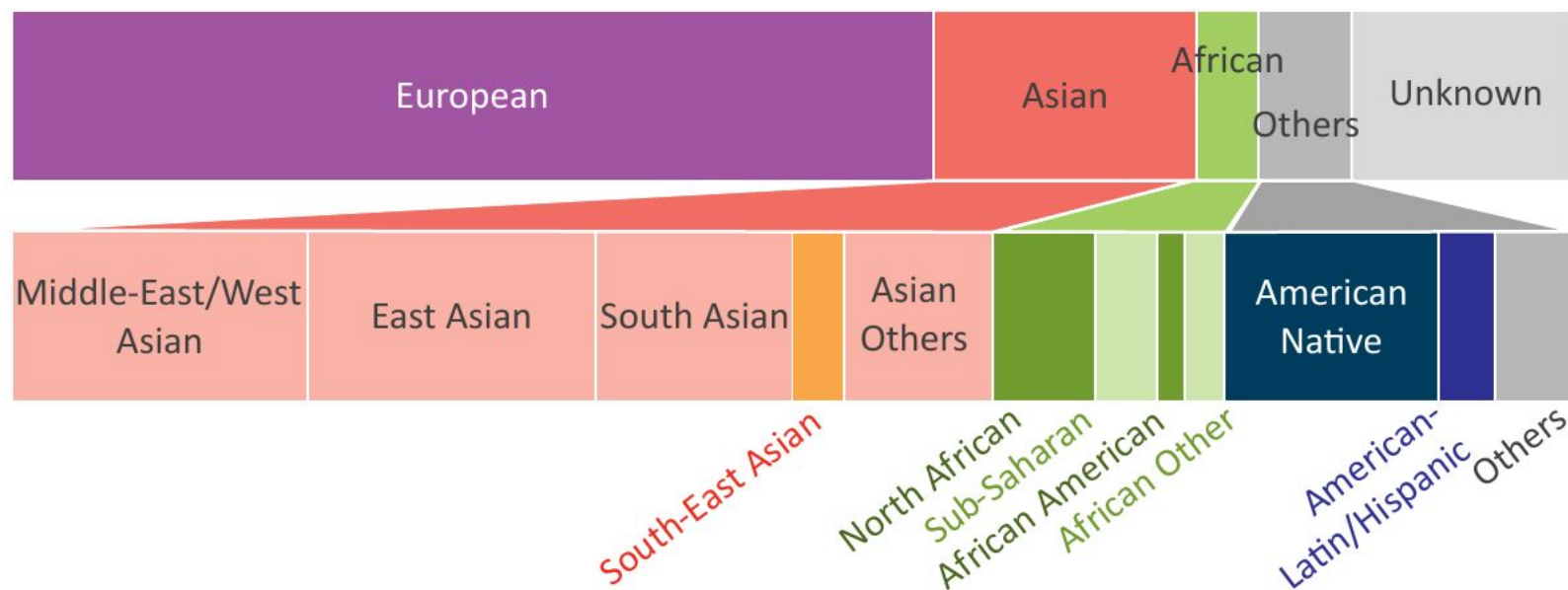
Scan me:



Ancestry and reference-data composition

GMDB-FAIR: 10,980 images / 8,346 patients / 581 disorders

Known ancestry: European 67%, Asian 19%, African 7%, Others 7%



Source graphic includes unknown ancestry. Reported labels are not genetic measurements.

Disease mix changes ancestry comparisons

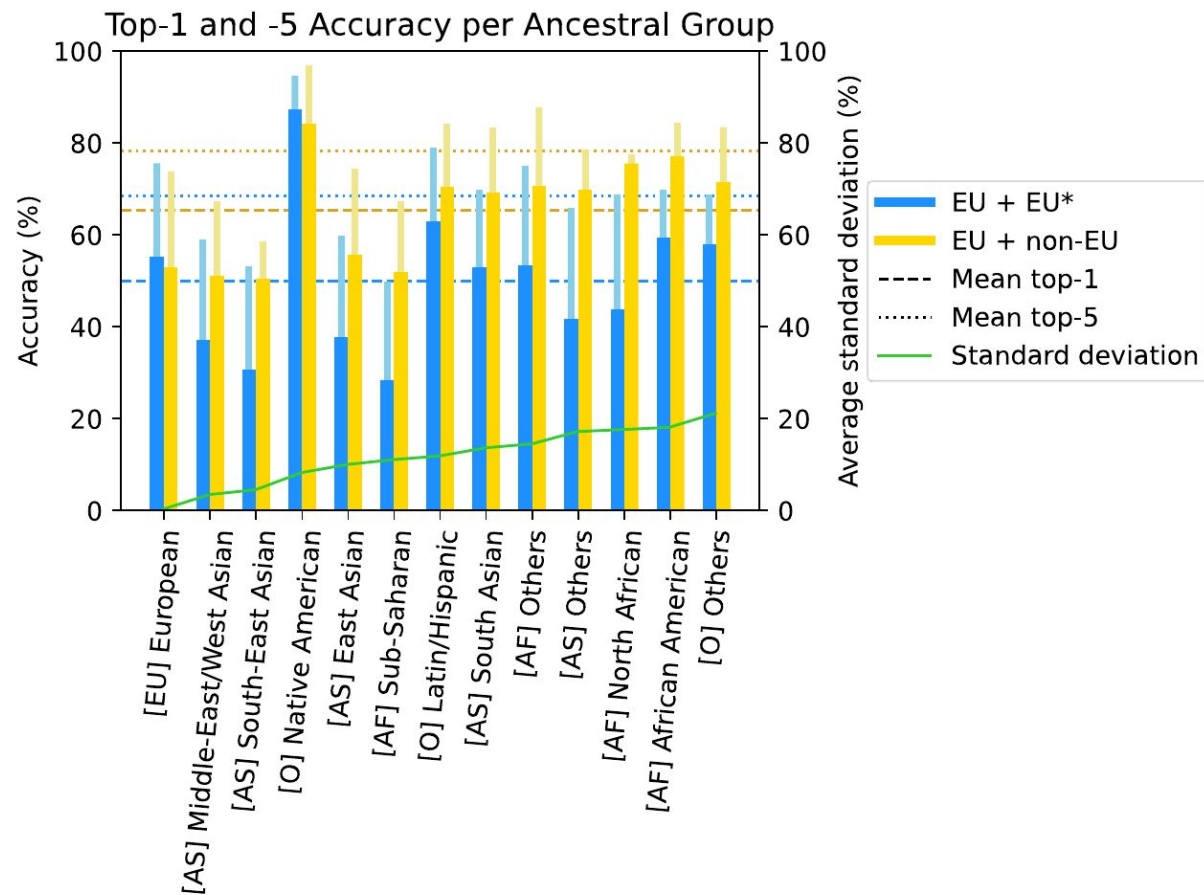
Top-1 accuracy: pooled versus shared-disorder comparison

Comparison	African	European
All tested disorders	62.07% 29 images	55.64% 523 images
Six shared disorders	64.29% 14 images	81.25% 32 images

Different disease mixes can conceal gaps. Shared-disorder samples remain small.

Diversity in the training data

a



+17.07 points

Non-European average
top-1 improvement

European top-1:
-2.26 points in
the broad comparison



Follow this preprint

GestaltMatcher Database - A global reference for facial phenotypic variability in rare human diseases

Hellen Lesmann, Alexander Hustinx, Shahida Moosa, Hannah Klinkhammer, Elaine Marchi, Pilar Caro, Ibrahim M. Abdelrazek, Jean Tori Pantel, Merle ten Hagen, Meow-Keong Thong, Rifhan Azwani Binti Mazlan, Sok Kun Tae, Tom Kamphans, Wolfgang Meiswinkel, Jing-Mei Li, Behnam Javanmardi, Alexej Knaus, Annette Uwineza, Cordula Knopp, Tinatin Tkemaladze, Miriam Elbracht, Larissa Mattern, Rami Abou Jamra, Clara Velmans, Vincent Strehlow, Maureen Jacob, Angela Peron, Cristina Dias, Beatriz Carvalho Nunes, Thainá Vilella, Isabel Furquim Pinheiro, Chong Ae Kim, Maria Isabel Melaragno, Hannah Weiland, Sophia Kaptain, Karolina Chwiałkowska, Mirosław Kwasniewski, Ramy Saad, Sarah Wiethoff, Himanshu Goel, Clara Tang, Anna Hau, Tahsin Stefan Barakat, Przemysław Panek, Amira Nabil, Julia Suh, Frederik Braun, Israel Gomy, Luisa Averdunk, Ekanem Ekure, Gaber Bergant, Borut Peterlin, Claudio Graziano, Nagwa Gaboon, Moisés Fiesco-Roa, Alessandro Mauro Spinelli, Nina-Maria Wilpert, Prasit Phowthongkum, Nergis Güzel, Tobias B. Haack, Rana Bitar, Andreas Tzschach, Agustí Rodríguez-Palmero, Theresa Brunet, Sabine Rudnik-Schöneborn, Silvina Noemi Contreras-Capetillo, Ava Oberlack, Carole Samango-Sprouse, Teresa Sadeghin, Margaret Olaya, Konrad Platzer, Artem Borovikov, Franziska Schnabel, Lara Heuft, Vera Herrmann, Renske Oegema, Nour Elkhateeb, Sheetal Kumar, Katalin Komlosi, Khoushoua Mohamed, Silvia Kalantari, Fabio Sirchia, Antonio F. Martinez-Monseny, Matthias Höller, Louiza Toutouna, Amal Mohamed, Amaia Lasa-Aranzasti, John A. Sayer, Nadja Ehmke, Magdalena Danyel, Henrike Sczakiel, Sarina Schwartzmann, Felix Boschann, Max Zhao, Ronja Adam, Lara Einicke, Denise Horn, Kee Seang Chew, Choy Chen Kam, Miray Karakoyun, Ben Pode-Shakked, Aviva Eliyahu, Rachel Rock, Teresa Carrion, Odelia Choring, Yuri A. Zarate, Marcelo Martinez Conti, Mert Karakaya, Moon Ley Tung, Bharatendu Chandra, Arjan Bouman, Aime Lumaka, Naveed Wasif, Marwan Shinawi, Patrick R. Blackburn, Tianyun Wang, Tim Niehues, Axel Schmidt, Regina Rita Roth, Dagmar Wiecezorek, Ping Hu, Rebekah L. Waikel, Suzanna E. Ledgister Hanchard, Gehad Elmakawy, Sylvia Safwat, Frédéric Ebstein, Elke Krüger, Sébastien Küry, Stéphane Béziau, Annabelle Arlt, Eric Olinger, Felix Marbach, Dong Li, Lucie Dupuis, Roberto Mendoza-Londono, Sofia Douzgou Houge, Denisa Weis, Brian Hon-Yin Chung, Christopher C.Y. Mak, Hülya Kayserili, Nursel Elcioglu, Ayca Aykut, Peli Özlem Şimşek-Kiper, Nina Bögershausen, Bernd Wollnik, Heidi Beate Bentzen, Ingo Kurth, Christian Netzer, Aleksandra Jezela-Stanek, Koen Devriendt, Karen W. Gripp, Martin Mücke, Alain Verloes, Christian P. Schaaf, Christoffer Nellåker, Benjamin D. Solomon, Markus M. Nöthen, Ebtesam Abdalla, Gholson J. Lyon, Peter M. Krawitz, Tzung-Chien Hsieh

doi: <https://doi.org/10.1101/2023.06.06.23290887>

Integrating LLM with facial image and clinical text



Please diagnose this patient.

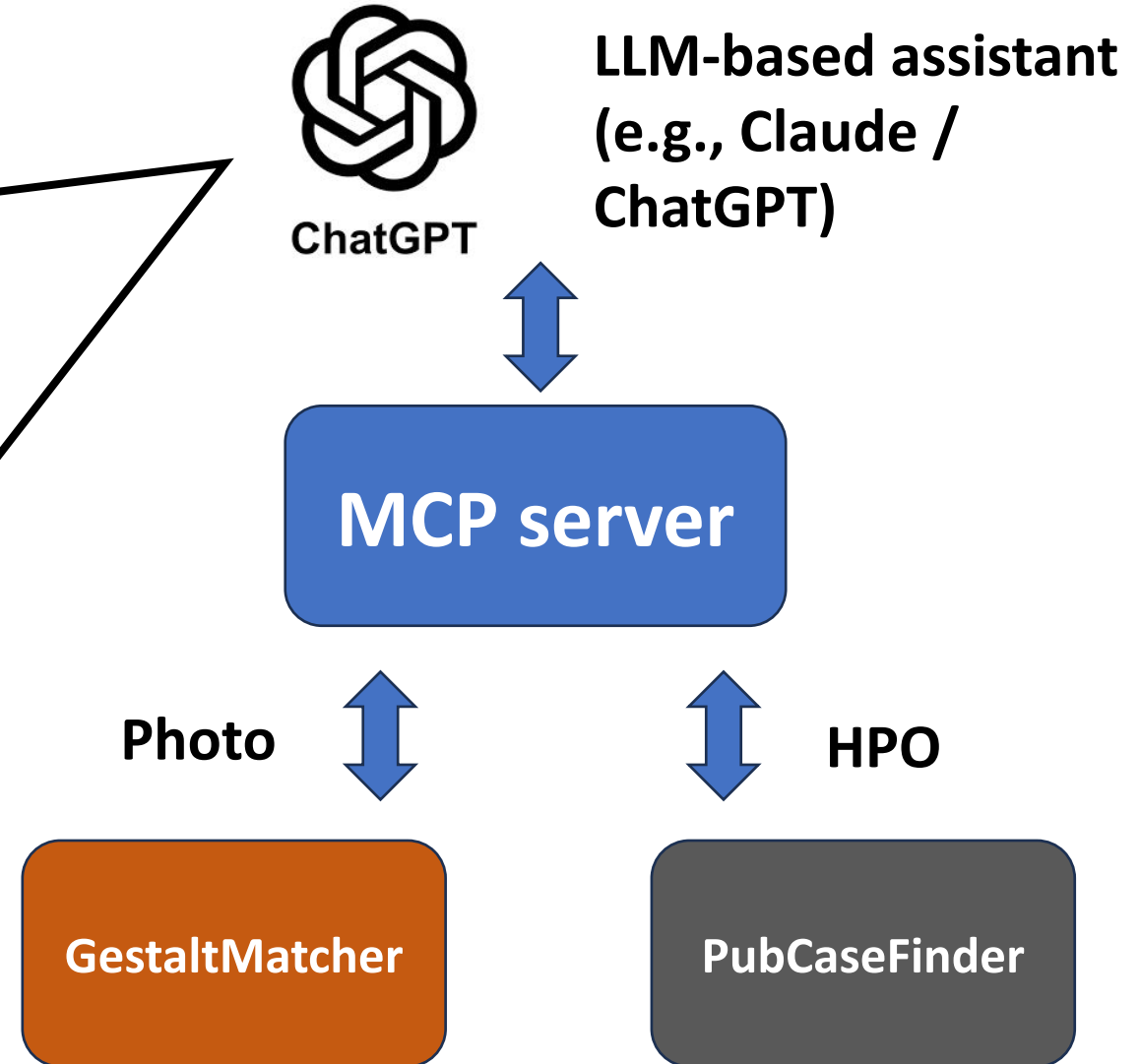
Patient Summary Demographics

- * Sex: Female
- * Age: 4 years
- * Ethnicity: American – Latin/Hispanic

Clinical Presentation The patient presents with multiple craniofacial and systemic abnormalities:

- * Craniofacial features:
 - * Abnormality of the oral cavity
 - * Abnormality of the eye
 - * Abnormality of the head

...



Integrating LLM with facial image and clinical text



I'll analyze this case using the available rare disease diagnostic tools to help identify potential diagnoses based on the clinical features and facial photograph. First, let me use the facial analysis tool to get additional diagnostic insights from **GestaltMatcher predict endpoint**

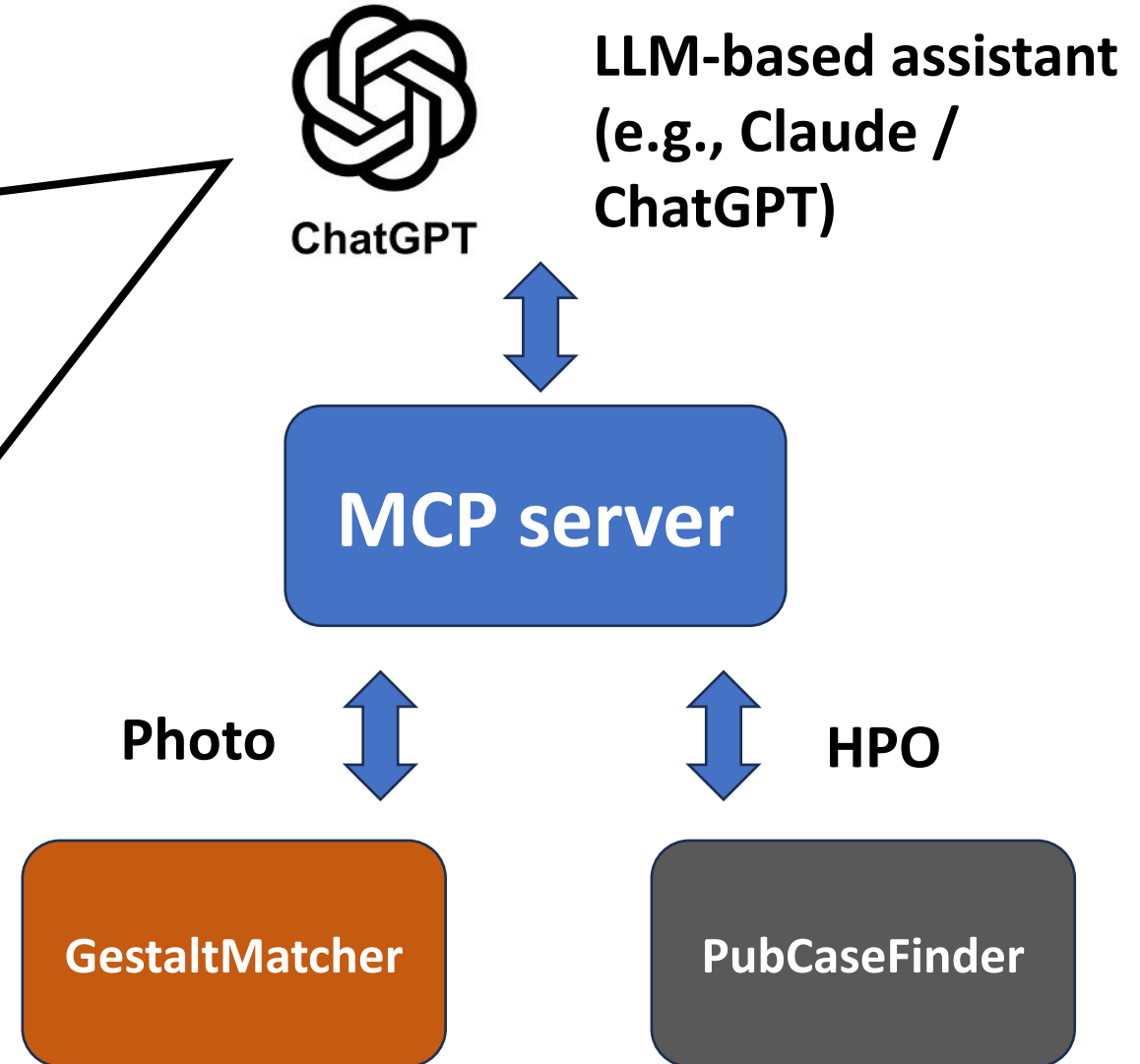
...

Let me search for conditions using individual HPO terms. Let me start with micrognathia (small jaw):**PubCaseFinder endpoint**

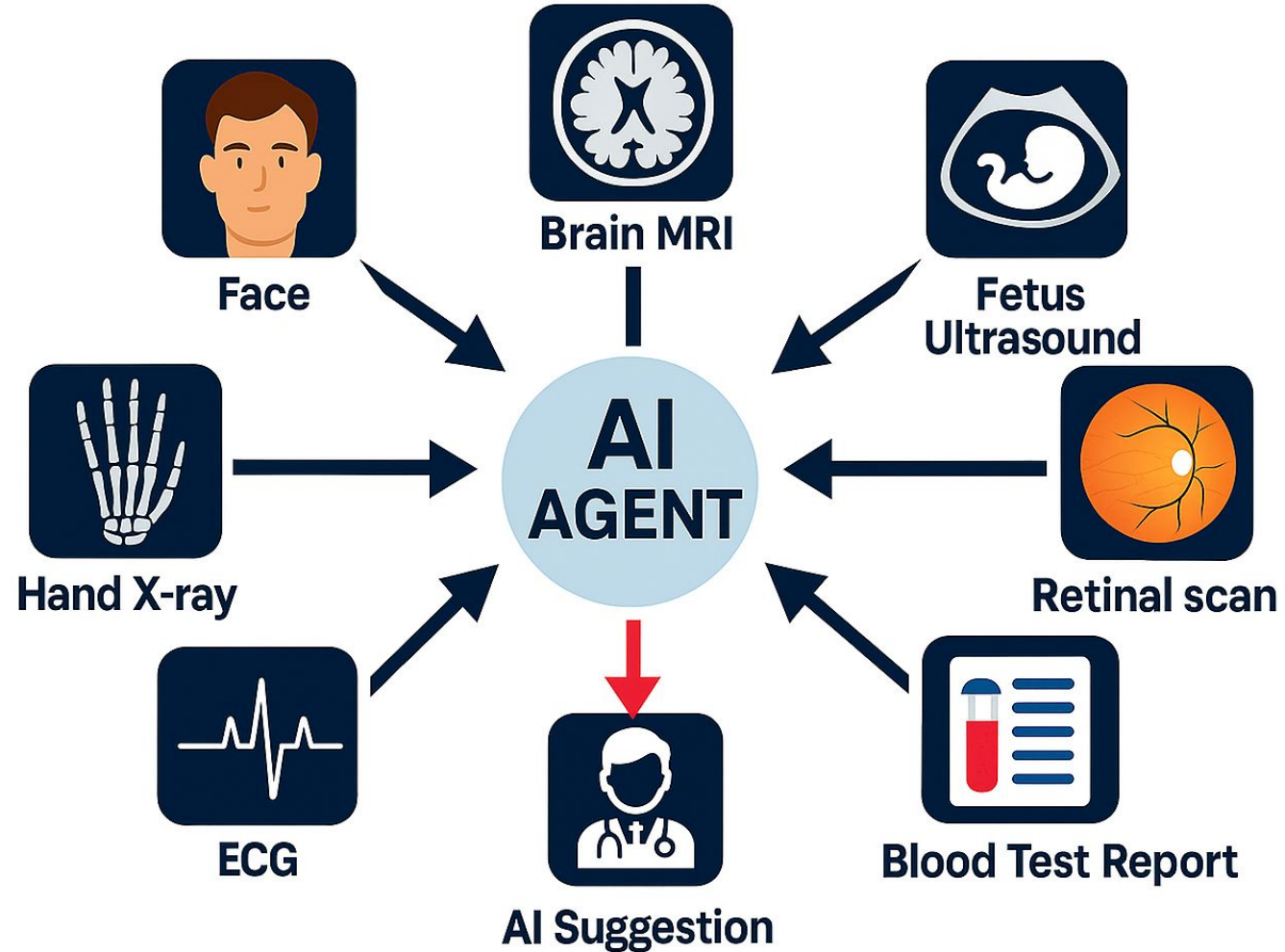
...

Diagnostic Considerations

1. **Cornelia de Lange Syndrome (CdLS)** - Most Likely OMIM:122470 - This appears to be the strongest diagnostic possibility based on the clinical overlap: Matching features:



Agentic AI integrating multi-modality analysis



Artificial Intelligence leads to loss of skills (deskilling)



THE LANCET

PERSPECTIVES | Digital medicine · Volume 406, Issue 10513, P1719, October 18,

2025

[Download Full Issue](#)

Preserving clinical skills in the age of AI assistance

[Tyler M Berzin](#)^a · [Eric J Topol](#)^b

THE LANCET

Gastroenterology & Hepatology

ARTICLES · Volume 10, Issue 10, P896-903, October 2025

[Download Full Issue](#)

Endoscopist deskilling risk after exposure to artificial intelligence in colonoscopy: a multicentre, observational study

[Krzysztof Budzyń, MD](#)^{a,b} · [Marcin Romańczyk, MD](#)^{a,b} · [Diana Kitala, PhD](#)^c · [Paweł Kołodziej, MD](#)^d · [Marek Bugajski, MD](#)^e · [Hans O Adami, MD](#)^{f,g} · [Johannes Blom, MD](#)^{h,i} · [Marek Buszkiewicz, MD](#)^j · [Natalie Halvorsen, MD](#)^f · [Prof Cesare Hassan, MD](#)^{k,l} · [Tomasz Romańczyk, MD](#)^{a,b} · [Prof Øyvind Holme, MD](#)^{m,n} · [Krzysztof Jarus, MD](#)^o · [Shona Fielding, PhD](#)^p · [Melina Kunar, PhD](#)^q · [Prof Maria Pellise, MD](#)^{r,s} · [Nastazja Pilonis, MD](#)^{t,u} · [Prof Michał Filip Kamiński, MD](#)^{f,v,w} · [Prof Mette Kalager, MD](#)^f · [Prof Michael Bretthauer, MD](#)^f · [Prof Yuichi Mori, MD](#)^{f,x} [Show less](#)

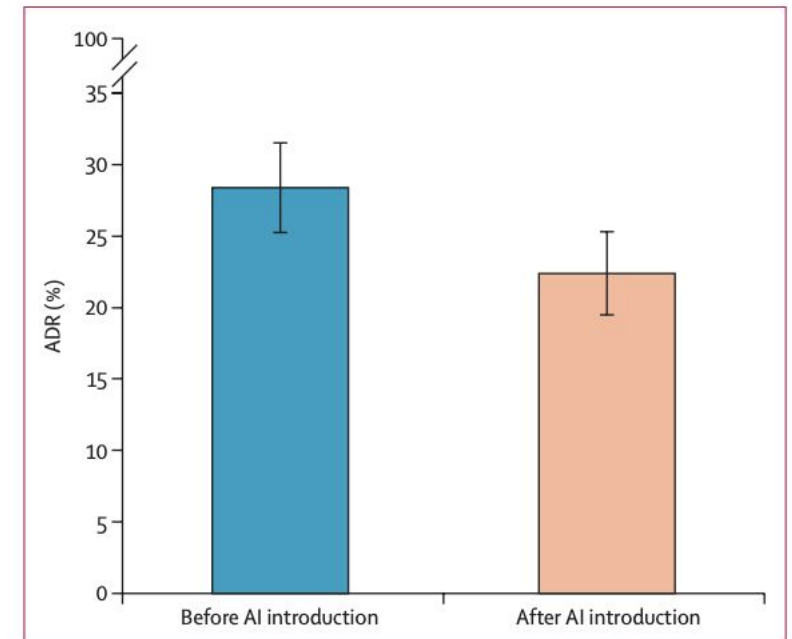
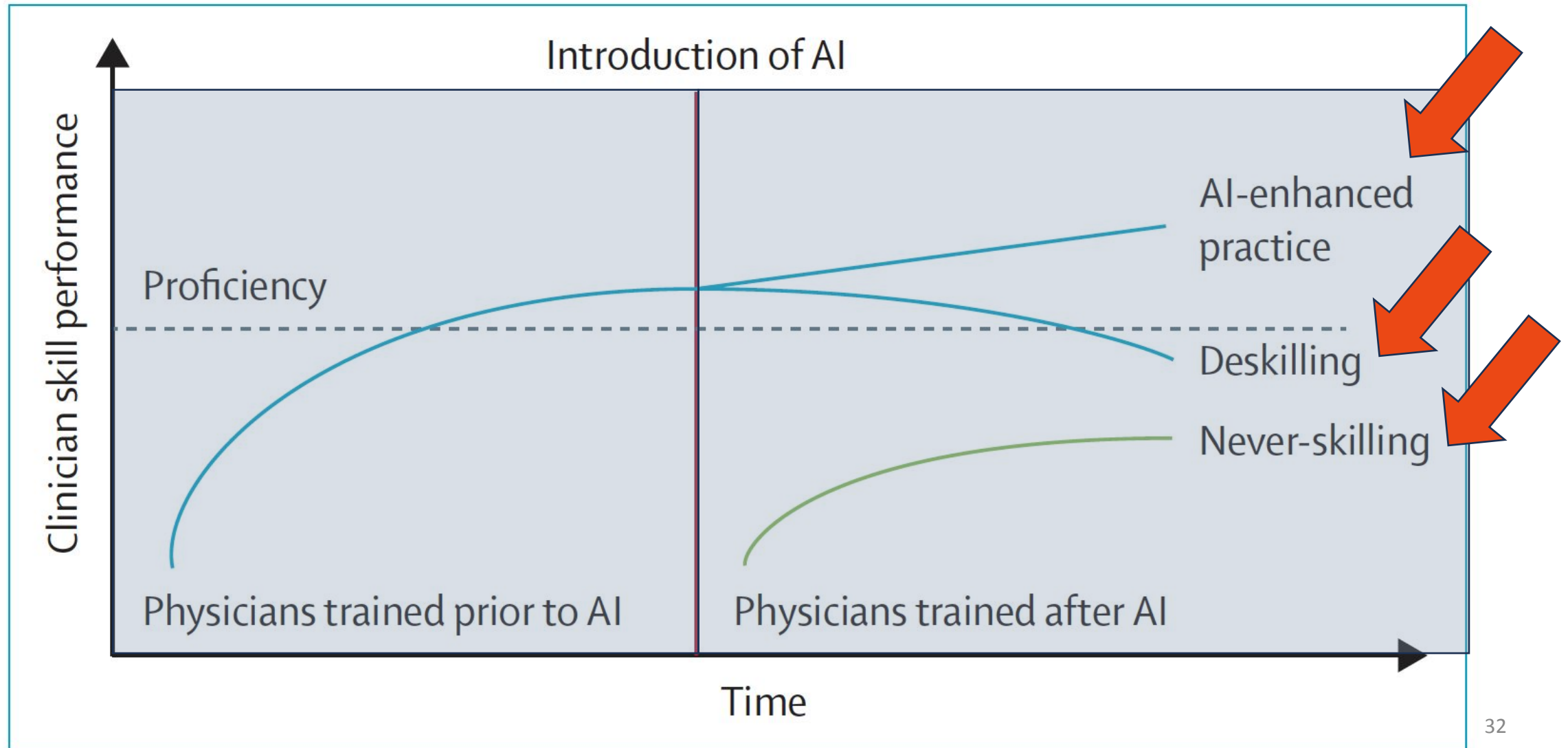


Figure 1: Change in ADR with standard, non-AI assisted colonoscopy before and after introduction of AI for polyp detection

The error bars represent 95% CIs. ADR=adenoma detection rate. AI=artificial intelligence.

Potential effects of AI on clinical skills over time

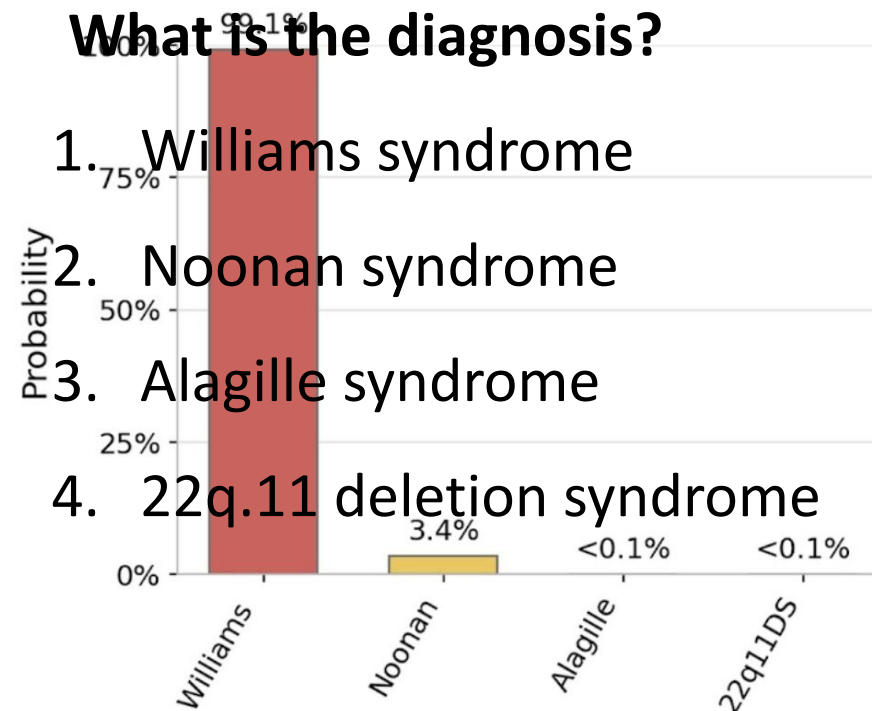


Will AI help clinicians make the diagnosis?



Prediction probability by AI

What is the diagnosis?



1. Williams syndrome
2. Noonan syndrome
3. Alagille syndrome
4. 22q.11 deletion syndrome

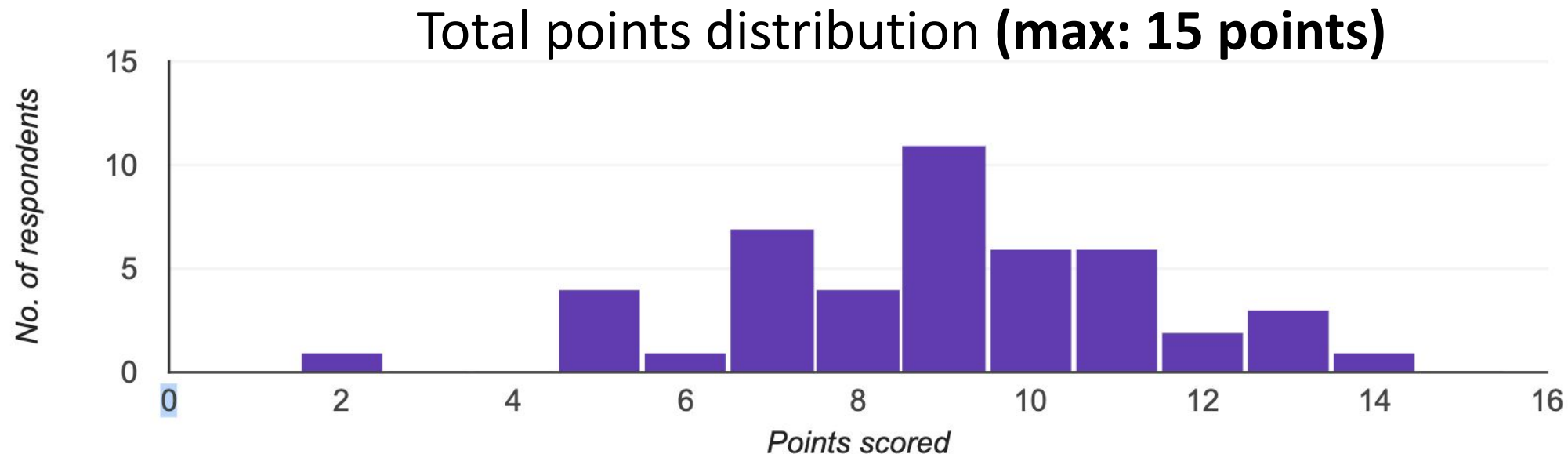
Will AI change your decision after seeing AI?

1. Williams syndrome
2. Noonan syndrome
3. Alagille syndrome
4. 22q.11 deletion syndrome

Clinicians can be influenced by AI predictions



In our first group, **39 of 40 clinicians changed from a correct to an incorrect answer** after seeing an incorrect AI suggestion.



Acknowledgements



IGSB

Prof. Peter Krawitz
Tzung-Chien Hsieh
Alexander Hustinx
Alexej Knaus
Tom Kamphans
Hellen Lesmann
Jing-Mei Li
Behnam Javanmardi
Wolfgang Meiswinkel
and all the colleagues

Thank you!



T-NAMSE

Prof. Christian Schaaf
Prof. Regina Betz
Theresa Brunet
Hartmut Engels
Sophia Peters
Axel Schmidt
Matias Wagner
Tobias Haack



Charité Berlin

Martin Mensah
Jean Tori Pantel
Stanislav Rosnev



CHU de Nantes
Sébastien KÜRY

Follow us

 **@GestaltMatcher**



Stellenbosch

Prof. Shahida Moosa
Kimberly Christine
Coetzer



thsieh@uni-bonn.de

