Automatic Detection of Chromosomal Abnormalities

Auriane Mazuir

Supervised by Étienne Birmelé, Augustin Chevallier and Gatien Ricotier

University of Strasbourg — B2A Laboratories

Introduction / State of the Art

Karyotyping consists in preparing cells arrested in metaphase, staining chromosomes using banding techniques (G, R, Q), then photographing and arranging the 46 human chromosomes (22 autosomal pairs + XX/XY) for visual inspection [1].

It allows the detection of abnormalities in **number** (e.g. trisomy 21, Turner syndrome, Klinefelter syndrome) or **structure** (deletions, duplications, inversions, translocations), and remains a cornerstone of prenatal, postnatal and hematological diagnostics.

However, the analysis is **time-consuming**, **highly dependent on the cytogeneticist's expertise**, and sensitive to preparation quality, leading to **inter-operator variability** [2]. To improve speed, robustness and **interpretability**, we propose to represent each chromosome by a **mathematical density profile** (*functional data analysis*), enabling direct curve-based comparisons. In parallel, AI-based approaches (segmentation, classification) exist but remain heterogeneous and poorly standardized [3].

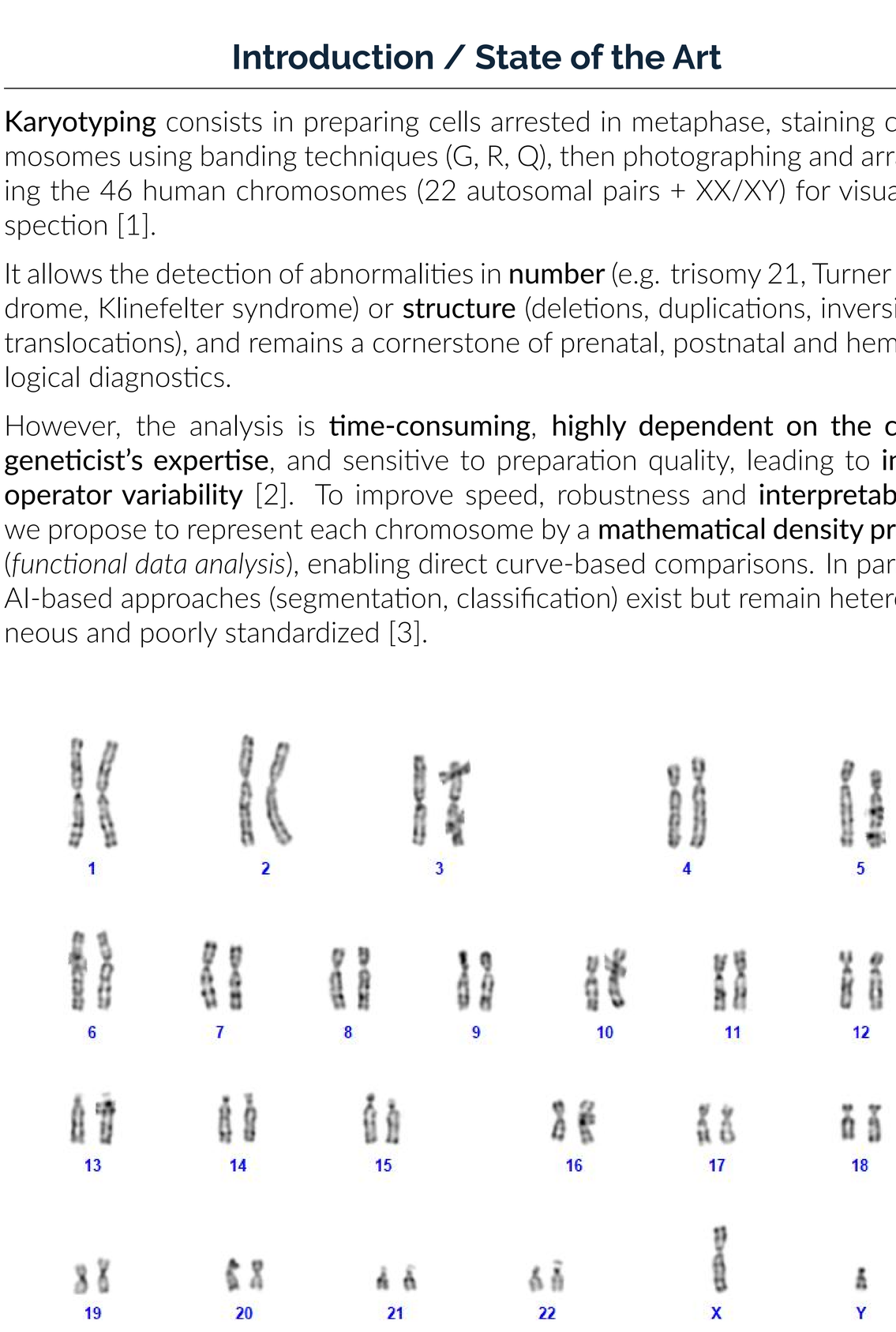

Figure 1. Human karyotype without notable abnormality (46,XY).

Problem Statement

Given that manual cytogenetic analysis is **time-consuming** and **expert-dependent**, how can we design an *image* → *functional profile* → *comparison* pipeline that allows:

- automatic extraction of **straight, isolated chromosomes**,
- identification and handling of **chromosome overlaps**,
- construction of **normalized density profiles** on $[0 - -100]$,
- comparison using **robust and interpretable distances**,
- and the computation of a **chromosome-level normality score**, aggregated over multiple karyotypes from the same patient.

Data

Sources. Binary files `Car.dat` (Genikon/Excilone) containing karyotype images.

Preprocessing and outputs.

- **Chromosome-wise extraction** (1–22, X, Y) and storage as `PNG`.
- **Tabular dataset**: image identifier, chromosomal class, vector $(X_1, \ldots, X_{100})$ (density profile interpolated on $[0, 1]$).

Current Progress: Straight Chromosome Extraction and Density Profiles

Implemented pipeline.

1. **Extraction** from `Car.dat` binary files, directly retrieving the **straightened version** of each chromosome, classified as 1–22, X, Y.
2. **Density profile computation**: obtained by extracting the **medial axis** of each chromosome and measuring the **intensity along this axis** to generate a one-dimensional density signal. This step is performed using the `Banding-Pattern-Extraction` package [4], followed by **normalization** and **uniform interpolation** (100 points on $[0, 1]$).
3. **Structured database** ready for functional statistics (comparison, clustering, anomaly detection).

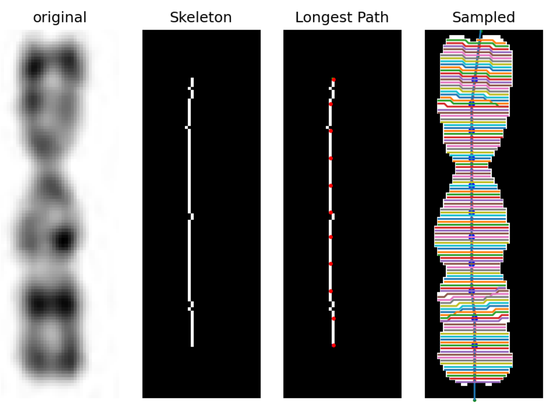

Figure 2. Extraction of the central axis of a chromosome.

Objectives

1. Robust comparison and alignment Define suitable distances between profiles: L^2 norm (global differences), correlation-based distance $(1 - \rho)$, and DTW (tolerance to local shifts). Introduce a **penalization of excessive compressions/dilations** and **automatic masking** of poorly informative regions.

2. Signal enhancement Correct **stain diffusion effects** using deconvolution methods (Wiener, Richardson–Lucy) under smoothness and non-negativity constraints. Harmonize profiles across preparations by handling **contrast inversions** depending on staining techniques and normalizing intensities for improved comparability.

3. Functional statistics and morphometry Leverage functional data analysis tools (curve depth, robust FPCA, high-density regions) to identify atypical profiles. Complement this approach with simple **morphometric indicators**: length-to-width ratio, pixel width along the axis, and integrate them jointly with density profiles.

4. Overlap detection Implement an automatic step to **detect overlapping or touching chromosomes**, allowing them to be **flagged as abnormal cases** and directly integrated into classification.

5. Towards a normality score Construct a **global chromosome-level normality score** combining functional and morphometric indicators. This score will be computed across **multiple karyotypes from the same patient** to assess intra-patient consistency and improve sensitivity to structural abnormalities and mosaicism.

References

[1] McGowan-Jordan J., Simons A., Schmid M. (2016). *ISCN 2016: International System for Human Cytogenomic Nomenclature*. Karger.

[2] Van Campen J. (2022). *Karyotype*. NHS England.

[3] Menaka D., Ganesh Vaidyanathan S. (2022). Chromenet: a CNN architecture with comparison of optimizers for classification of human chromosome images. *Multidimensional Systems and Signal Processing*, 33:747–768.

[4] Lukasuz (2024). *Banding-Pattern-Extraction* [source code]. GitHub.