



Article review: MCPS and snipar

ROBERTO OLVERA-HERNANDEZ

*Centre for Genomic Sciences (CCG),
National Autonomous University of Mexico (UNAM)*

Biweekly meeting, February 18, 2025



Overview

Over **150,000 participants** were recruited in two districts between **1998 and 2004**.

- ▶ Baseline questionnaire.
- ▶ Blood samples.
- ▶ Physical measurements.
- ▶ Linkage to mortality.

COHORT PROFILE

Cohort Profile: The Mexico City Prospective Study

Roberto Tapia-Conyer,¹ Pablo Kuri-Morales,² Jesús Alegre-Díaz,² Gary Whitlock,^{3*}
Jonathan Emberson,³ Sarah Clark,² Richard Peto³ and Rory Collins³

(b) Mexico City



Figure: Map showing the location of the MCPS districts (Tapia-Conyer et al. 2006).



Baseline data

Socio-demographic

- ▶ Age and sex
- ▶ Area of residence
- ▶ Marital status
- ▶ Educational achievement
- ▶ Occupation
- ▶ Income
- ▶ Health service provider



Baseline data

Socio-demographic

- ▶ Age and sex
- ▶ Area of residence
- ▶ Marital status
- ▶ Educational achievement
- ▶ Occupation
- ▶ Income
- ▶ Health service provider

Lifestyle characteristics

- ▶ Diet (fruit/vegetables, fried food, types of oil)
- ▶ Smoking and alcohol
- ▶ Physical activity
- ▶ Sleep duration



Baseline data

Socio-demographic

- ▶ Age and sex
- ▶ Area of residence
- ▶ Marital status
- ▶ Educational achievement
- ▶ Occupation
- ▶ Income
- ▶ Health service provider

Reproductive history (women)

- ▶ Menopausal status
- ▶ Hysterectomy
- ▶ Oopherectomy
- ▶ HRT
- ▶ Contraceptive use
- ▶ Pregnancy (age and number)

Lifestyle characteristics

- ▶ Diet (fruit/vegetables, fried food, types of oil)
- ▶ Smoking and alcohol
- ▶ Physical activity
- ▶ Sleep duration



Baseline data

Socio-demographic

- ▶ Age and sex
- ▶ Area of residence
- ▶ Marital status
- ▶ Educational achievement
- ▶ Occupation
- ▶ Income
- ▶ Health service provider

Reproductive history (women)

- ▶ Menopausal status
- ▶ Hysterectomy
- ▶ Oopherectomy
- ▶ HRT
- ▶ Contraceptive use
- ▶ Pregnancy (age and number)

Lifestyle characteristics

- ▶ Diet (fruit/vegetables, fried food, types of oil)
- ▶ Smoking and alcohol
- ▶ Physical activity
- ▶ Sleep duration

Physical measurements

- ▶ Height
- ▶ Weight
- ▶ Waist and hip circumference
- ▶ Systolic and diastolic blood pressure



Baseline data

Socio-demographic

- ▶ Age and sex
- ▶ Area of residence
- ▶ Marital status
- ▶ Educational achievement
- ▶ Occupation
- ▶ Income
- ▶ Health service provider

Reproductive history (women)

- ▶ Menopausal status
- ▶ Hysterectomy
- ▶ Oopherectomy
- ▶ HRT
- ▶ Contraceptive use
- ▶ Pregnancy (age and number)

Blood samples

- ▶ Plasma & buffy coat
- ▶ HbA1c and other essays
- ▶ NMR metabolomics (e.g. fatty acids, cholines, lipoprotein subclasses, etc.)

Lifestyle characteristics

- ▶ Diet (fruit/vegetables, fried food, types of oil)
- ▶ Smoking and alcohol
- ▶ Physical activity
- ▶ Sleep duration

Physical measurements

- ▶ Height
- ▶ Weight
- ▶ Waist and hip circumference
- ▶ Systolic and diastolic blood pressure



Baseline data

Socio-demographic

- ▶ Age and sex
- ▶ Area of residence
- ▶ Marital status
- ▶ Educational achievement
- ▶ Occupation
- ▶ Income
- ▶ Health service provider

Lifestyle characteristics

- ▶ Diet (fruit/vegetables, fried food, types of oil)
- ▶ Smoking and alcohol
- ▶ Physical activity
- ▶ Sleep duration

Reproductive history (women)

- ▶ Menopausal status
- ▶ Hysterectomy
- ▶ Oopherectomy
- ▶ HRT
- ▶ Contraceptive use
- ▶ Pregnancy (age and number)

Physical measurements

- ▶ Height
- ▶ Weight
- ▶ Waist and hip circumference
- ▶ Systolic and diastolic blood pressure

Blood samples

- ▶ Plasma & buffy coat
- ▶ HbA1c and other essays
- ▶ NMR metabolomics (e.g. fatty acids, cholines, lipoprotein subclasses, etc.)

Prior diseases and medications

Participants were asked if they had ever been diagnosed with any of the listed diseases (binary: Yes or No).



Genetic datasets

Genetic datasets were added later by Ziyatdinov et al. (2023), making it one of the **largest** studies for **non-european** populations.



Genetic datasets

Genetic datasets were added later by Ziyatdinov et al. (2023), making it one of the **largest** studies for **non-european** populations.

Genome-Wide Genotyping

- ▶ Illumina — GSAv2 chip array
- ▶ 138,511 individuals



Genetic datasets

Genetic datasets were added later by Ziyatdinov et al. (2023), making it one of the **largest** studies for **non-european** populations.

Genome-Wide Genotyping

- ▶ Illumina — GSAv2 chip array
- ▶ 138,511 individuals

Exome Sequencing (WES)

- ▶ $n = 141,046$ individuals
- Variants:**
- ▶ *Total:* 9.3 million.
 - ▶ *Coding regions:* 4.0 million in 19,110 genes.
 - ▶ *Unique MCPS:* 1.4 million.



Genetic datasets

Genetic datasets were added later by Ziyatdinov et al. (2023), making it one of the **largest** studies for **non-european** populations.

Genome-Wide Genotyping

- ▶ Illumina — GSAv2 chip array
- ▶ 138,511 individuals

Exome Sequencing (WES)

- ▶ $n = 141,046$ individuals

Variants:

- ▶ *Total*: 9.3 million.
- ▶ *Coding regions*: 4.0 million in 19,110 genes.
- ▶ *Unique MCPS*: 1.4 million.

Whole-Genome Sequencing (WGS)

- ▶ $n = 9,950$ individuals

Variants:

- ▶ *Total*: 131.9 million.
- ▶ *Coding regions*: 1.5 million.
- ▶ *Unique MCPS*: 31.5 million.



Genetic datasets

Genetic datasets were added later by Ziyatdinov et al. (2023), making it one of the **largest** studies for **non-european** populations.

Genome-Wide Genotyping

- ▶ Illumina — GSAv2 chip array
- ▶ 138,511 individuals

Exome Sequencing (WES)

- ▶ $n = 141,046$ individuals

Variants:

- ▶ *Total*: 9.3 million.
- ▶ *Coding regions*: 4.0 million in 19,110 genes.
- ▶ *Unique MCPS*: 1.4 million.

Whole-Genome Sequencing (WGS)

- ▶ $n = 9,950$ individuals

Variants:

- ▶ *Total*: 131.9 million.
- ▶ *Coding regions*: 1.5 million.
- ▶ *Unique MCPS*: 31.5 million.

Both **WES** and **WGS** share **93.2%** of the variants, with an increment of **2.3%** on **WGS** data.



Family networks

Estimation

Relatedness was estimated through *identity-by-descent (IBD)* sharing.

About 71% of individuals have **at least one relative** present in the MCPS dataset.

- ▶ **Parent-Offspring (PO):** 31,597 relationships.
- ▶ **Sibling Pairs (FS):** 29,482 relationships.
- ▶ **Second Degree (2nd):** 47,080 relationships.
- ▶ **Third Degree (3rd):** 120,180 relationships.

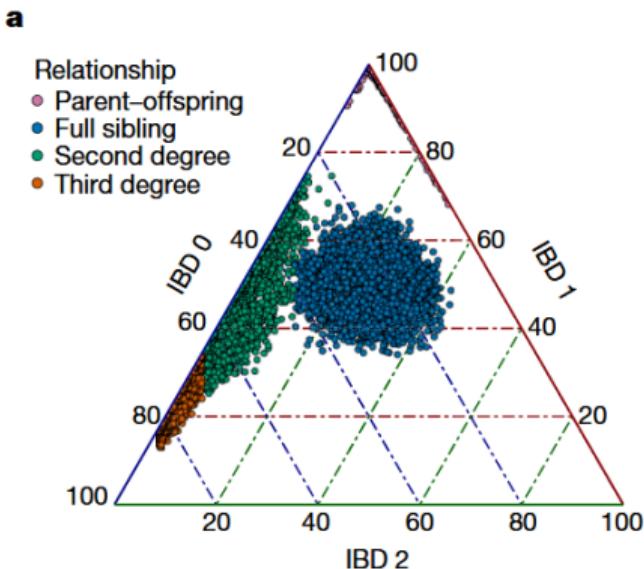


Figure: Percentage of the genome.



Family networks

The levels of *relatedness* were:

- ▶ much higher than those from the **UK Biobank (UKB)**.
- ▶ comparable with the **Geisinger Health Study (GHS)**—both MCPS and GHS recruited in *close proximity*.

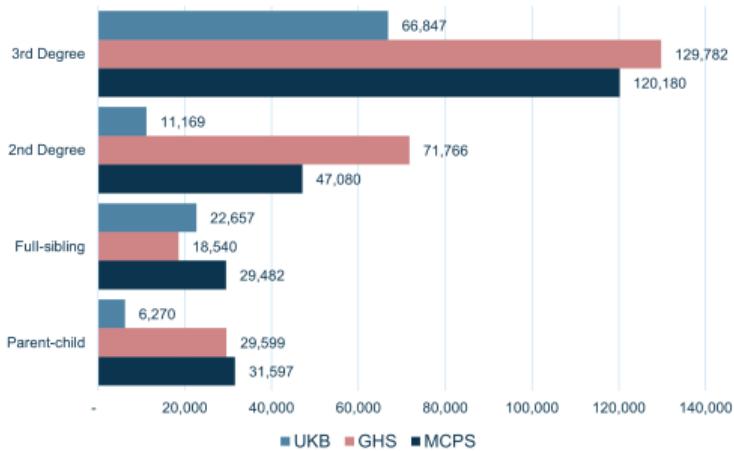


Figure: Comparison of network sizes in MCPS, UKB and GHS. Data extracted from Supplementary Table 25 (Ziyatdinov et al. 2023).



Family networks

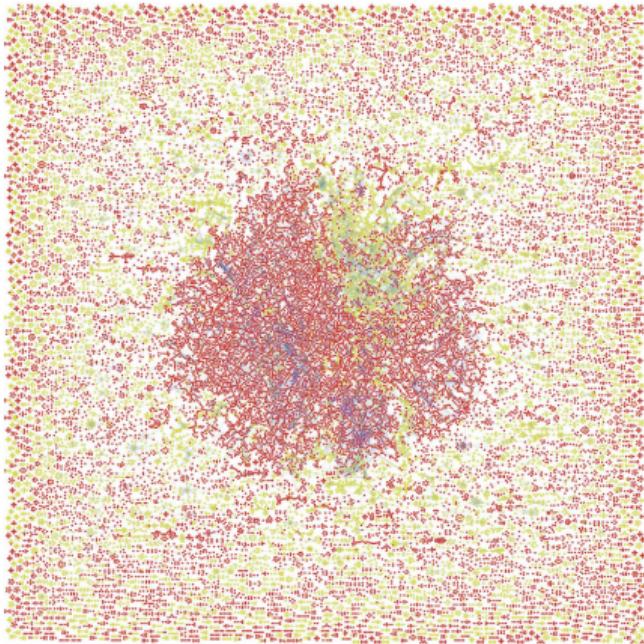


Figure: Graph of second-degree family networks of size four or greater (Ziyatdinov et al. 2023).



References



Tapia-Conyer, R., Kuri-Morales, P., Alegre-Díaz, J., Whitlock, G., Emberson, J., Clark, S., Peto, R., & Collins, R. (2006). *International Journal of Epidemiology*, 35(2), 243–249. <https://doi.org/10.1093/ije/dyl042>



Ziyatdinov, A., Torres, J., Alegre-Díaz, J., Backman, J., Mbatchou, J., Turner, M., Gaynor, S. M., Joseph, T., Zou, Y., Liu, D., Wade, R., Staples, J., Panea, R., Popov, A., Bai, X., Balasubramanian, S., Habegger, L., Lanche, R., Lopez, A., ... Tapia-Conyer, R. (2023). *Nature*, 622(7984), 784–793. <https://doi.org/10.1038/s41586-023-06595-3>