



# 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,<sup>1</sup> Pablo Kuri-Morales,<sup>2</sup> Jesús Alegre-Díaz,<sup>2</sup> Gary Whitlock,<sup>3\*</sup>  
Jonathan Emberson,<sup>3</sup> Sarah Clark,<sup>2</sup> Richard Peto<sup>3</sup> and Rory Collins<sup>3</sup>

(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

## 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

## 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.

## Article

# Genotyping, sequencing and analysis of 140,000 adults from Mexico City

<https://doi.org/10.1038/s41586-023-06595-3>  
Received: 27 June 2022  
Accepted: 31 August 2023  
Published online: 11 October 2023  
Open access  
 Check for updates

Andrey Ziyatdinov<sup>1,2\*</sup>, Jason Torres<sup>1,3,4,5,6,7</sup>, Jesús Alegre-Díaz<sup>4,8</sup>, Joshua Backman<sup>9</sup>, Joelle Mbatchou<sup>1</sup>, Michael Turner<sup>1,3</sup>, Sheila M. Gaynor<sup>1</sup>, Tyler Joseph<sup>1</sup>, Yuxin Zou<sup>1</sup>, Daren Liu<sup>1</sup>, Rachel Wade<sup>1,3</sup>, Jeffrey Staples<sup>1</sup>, Razvan Panea<sup>1</sup>, Alex Popov<sup>1</sup>, Xiaodong Bai<sup>1</sup>, Suganthi Balasubramanian<sup>1</sup>, Lukas Habegger<sup>1</sup>, Rouet Lanche<sup>1</sup>, Alex Lopez<sup>1</sup>, Evan Maxwell<sup>1</sup>, Marcus Jones<sup>1</sup>, Humberto García-Ortiz<sup>1</sup>, Raul Ramírez-Reyes<sup>1</sup>, Rogelio Santacruz-Benítez<sup>1</sup>, Abhishek Nag<sup>1</sup>, Katherine R. Smith<sup>1</sup>, Amy Damask<sup>1</sup>, Nan Lin<sup>1</sup>, Charles Paufling<sup>1</sup>, Mark Reppell<sup>1</sup>, Sebastian Zöllner<sup>1</sup>, Eric Jorgenson<sup>1</sup>, William Salerno<sup>1</sup>, Slavé Petrovski<sup>1</sup>, John Overton<sup>1</sup>, Jeffrey Reid<sup>1</sup>, Timothy A. Thornton<sup>1</sup>, Gonzalo Abecasis<sup>1</sup>, Jaime Berumen<sup>1</sup>, Lorena Orozco-Orozco<sup>1</sup>, Rory Collins<sup>2</sup>, Regeneron Genetics Center<sup>1</sup>, Mexico City Prospective Study<sup>1</sup>, Aris Baras<sup>1,8</sup>, Michael R. Hill<sup>1,3,6</sup>, Jonathan R. Emberson<sup>1,3,8</sup>, Jonathan Marchini<sup>1,3,9</sup>, Pablo Kuri-Morales<sup>10,11,12</sup> & Roberto Tapia-Conyer<sup>1,3,13</sup>



# Genetic datasets

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

## Sequences

- ▶ **Genome-Wide Genotyping**
  - ▶ Illumina — GSAv2 chip array
  - ▶  $n = 138,511$  individuals

### Article

## Genotyping, sequencing and analysis of 140,000 adults from Mexico City

<https://doi.org/10.1038/s41586-023-06595-3>

Received: 27 June 2022

Accepted: 31 August 2023

Published online: 11 October 2023

Open access

Check for updates

Andrey Ziyatdinov<sup>1,2\*</sup>, Jason Torres<sup>1,3,4,5,6</sup>, Jesús Alegre-Díaz<sup>4,6</sup>, Joshua Backman<sup>7</sup>, Joelle Mbatchou<sup>1</sup>, Michael Turner<sup>1,2</sup>, Sheila M. Gaynor<sup>8</sup>, Tyler Joseph<sup>9</sup>, Yuxin Zou<sup>1</sup>, Darren Liu<sup>1</sup>, Rachel Wade<sup>1,2</sup>, Jeffrey Staples<sup>9</sup>, Razvan Panca<sup>9</sup>, Alex Popov<sup>9</sup>, Xiaodong Bai<sup>1</sup>, Suganthi Balasubramanian<sup>1</sup>, Lukas Habegger<sup>1</sup>, Rouet Lanche<sup>1</sup>, Alex Lopez<sup>1</sup>, Evan Maxwell<sup>1</sup>, Marcus Jones<sup>1</sup>, Humberto García-Ortiz<sup>1</sup>, Raul Ramírez-Reyes<sup>1</sup>, Rogelio Santacruz-Benítez<sup>1</sup>, Abhishek Nag<sup>1</sup>, Katherine R. Smith<sup>1</sup>, Amy Damask<sup>1</sup>, Nan Lin<sup>1</sup>, Charles Paufling<sup>1</sup>, Mark Reppell<sup>1</sup>, Sebastian Zöllner<sup>9</sup>, Eric Jorgenson<sup>9</sup>, William Salerno<sup>9</sup>, Slavé Petrovski<sup>7</sup>, John Overton<sup>1</sup>, Jeffrey Reid<sup>7</sup>, Timothy A. Thornton<sup>1</sup>, Gonzalo Abecasis<sup>1</sup>, Jaime Berumen<sup>6</sup>, Lorena Orozco-Orozco<sup>2</sup>, Rory Collins<sup>2</sup>, Regeneron Genetics Center<sup>4</sup>, Mexico City Prospective Study<sup>1</sup>, Aris Baras<sup>1,9</sup>, Michael R. Hill<sup>1,3,6</sup>, Jonathan R. Emberson<sup>1,3,6</sup>, Jonathan Marchini<sup>1,3,6</sup>, Pablo Kuri-Morales<sup>1,3,6</sup> & Roberto Tapia-Conyer<sup>1,3,6</sup>



# Genetic datasets

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

## Sequences

- ▶ **Genome-Wide Genotyping**
  - ▶ Illumina — GSAv2 chip array
  - ▶  $n = 138,511$  individuals
- ▶ **Exome Sequencing (WES)**
  - ▶  $n = 141,046$  individuals

### Article

## Genotyping, sequencing and analysis of 140,000 adults from Mexico City

<https://doi.org/10.1038/s41586-023-06595-3>

Received: 27 June 2022

Accepted: 31 August 2023

Published online: 11 October 2023

Open access

Check for updates

Andrey Ziyatdinov<sup>1,2\*</sup>, Jason Torres<sup>1,3,4,5,6</sup>, Jesús Alegre-Díaz<sup>4,6</sup>, Joshua Backman<sup>7</sup>, Joelle Mbatchou<sup>1</sup>, Michael Turner<sup>1,2</sup>, Sheila M. Gaynor<sup>8</sup>, Tyler Joseph<sup>9</sup>, Yuxin Zou<sup>1</sup>, Daren Liu<sup>1</sup>, Rachel Wade<sup>1,3</sup>, Jeffrey Staples<sup>9</sup>, Razvan Panea<sup>9</sup>, Alex Popov<sup>1</sup>, Xiaodong Bai<sup>1</sup>, Suganthi Balasubramanian<sup>1</sup>, Lukas Habegger<sup>1</sup>, Rouet Lanche<sup>1</sup>, Alex Lopez<sup>1</sup>, Evan Maxwell<sup>1</sup>, Marcus Jones<sup>1</sup>, Humberto García-Ortiz<sup>1</sup>, Raul Ramírez-Reyes<sup>1</sup>, Rogelio Santacruz-Benítez<sup>1</sup>, Abhishek Nag<sup>1</sup>, Katherine R. Smith<sup>1</sup>, Amy Damask<sup>1</sup>, Nan Lin<sup>1</sup>, Charles Paufling<sup>1</sup>, Mark Reppell<sup>1</sup>, Sebastian Zöllner<sup>1</sup>, Eric Jorgenson<sup>1</sup>, William Salerno<sup>1</sup>, Slavé Petrovski<sup>1</sup>, John Overton<sup>1</sup>, Jeffrey Reid<sup>1</sup>, Timothy A. Thornton<sup>1</sup>, Gonzalo Abecasis<sup>1</sup>, Jaime Berumen<sup>6</sup>, Lorena Orozco-Orozco<sup>2</sup>, Rory Collins<sup>2</sup>, Regeneron Genetics Center<sup>1</sup>, Mexico City Prospective Study<sup>1</sup>, Aris Baras<sup>1,9</sup>, Michael R. Hill<sup>1,3,6</sup>, Jonathan R. Emberison<sup>1,3,6</sup>, Jonathan Marchini<sup>1,3,6</sup>, Pablo Kuri-Morales<sup>1,3,6</sup> & Roberto Tapia-Conyer<sup>1,3,6</sup>



# Genetic datasets

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

## Sequences

- ▶ **Genome-Wide Genotyping**
  - ▶ Illumina — GSAv2 chip array
  - ▶  $n = 138,511$  individuals
- ▶ **Exome Sequencing (WES)**
  - ▶  $n = 141,046$  individuals
- ▶ **Whole-Genome Sequencing (WGS)**
  - ▶  $n = 9,950$  individuals

### Article

## Genotyping, sequencing and analysis of 140,000 adults from Mexico City

<https://doi.org/10.1038/s41586-023-06595-3>

Received: 27 June 2022

Accepted: 31 August 2023

Published online: 11 October 2023

Open access

Check for updates

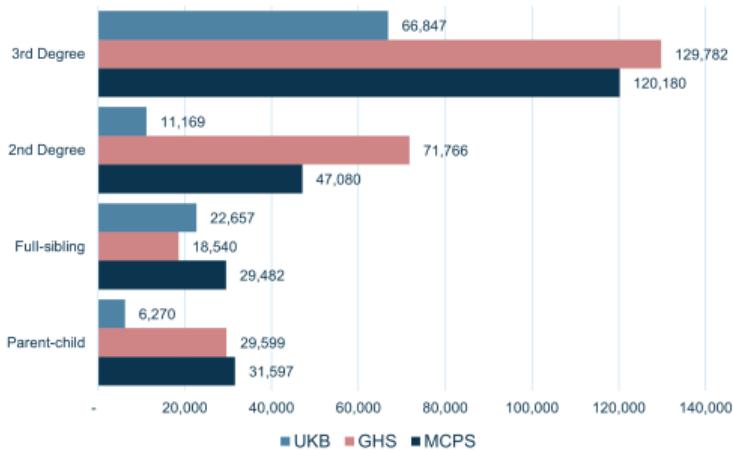
Andrey Ziyatdinov<sup>1,2\*</sup>, Jason Torres<sup>1,3,4,5,6</sup>, Jesús Alegre-Díaz<sup>4,6</sup>, Joshua Backman<sup>7</sup>, Joelle Mbatchou<sup>1</sup>, Michael Turner<sup>1,2</sup>, Sheila M. Gaynor<sup>8</sup>, Tyler Joseph<sup>9</sup>, Yuxin Zou<sup>1</sup>, Daren Liu<sup>1</sup>, Rachel Wade<sup>1,3</sup>, Jeffrey Staples<sup>1</sup>, Razvan Panca<sup>1</sup>, Alex Popov<sup>1</sup>, Xiaodong Bai<sup>1</sup>, Suganthi Balasubramanian<sup>1</sup>, Lukas Habegger<sup>1</sup>, Rouet Lanche<sup>1</sup>, Alex Lopez<sup>1</sup>, Evan Maxwell<sup>1</sup>, Marcus Jones<sup>1</sup>, Humberto García-Ortiz<sup>1</sup>, Raul Ramírez-Reyes<sup>1</sup>, Rogelio Santacruz-Benítez<sup>1</sup>, Abhishek Nag<sup>1</sup>, Katherine R. Smith<sup>1</sup>, Amy Damask<sup>1</sup>, Nan Lin<sup>1</sup>, Charles Paufling<sup>1</sup>, Mark Reppell<sup>1</sup>, Sebastian Zöllner<sup>1</sup>, Eric Jorgenson<sup>1</sup>, William Salerno<sup>1</sup>, Slavé Petrovski<sup>1</sup>, John Overton<sup>1</sup>, Jeffrey Reid<sup>1</sup>, Timothy A. Thornton<sup>1</sup>, Gonzalo Abecasis<sup>1</sup>, Jaime Berumen<sup>1</sup>, Lorena Orozco-Orozco<sup>1</sup>, Rory Collins<sup>1</sup>, Regeneron Genetics Center<sup>1</sup>, Mexico City Prospective Study<sup>1</sup>, Aris Baras<sup>1,9</sup>, Michael R. Hill<sup>1,3,6</sup>, Jonathan R. Emberison<sup>1,3,6</sup>, Jonathan Marchini<sup>1,3,6</sup>, Pablo Kuri-Morales<sup>1,3,6</sup> & Roberto Tapia-Conyer<sup>1,3,6</sup>



# 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

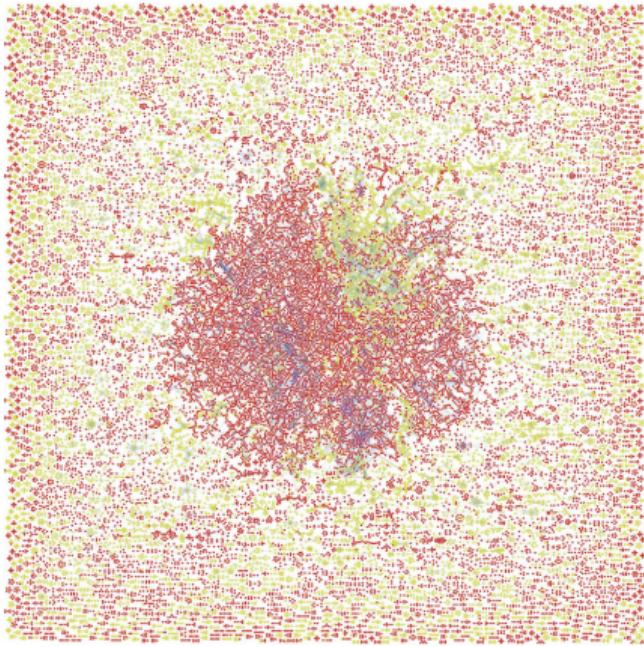


Figure: Graph of second-degree family networks of size four or greater (Ziyatdinov et al. 2023).



## Test for code (Python)

You can use the `string()` command to highlight a line of code.

```
1 def write_list(fid, x, level):
2     # Oh look, a comment!
3     ind = ' '*level
4     xs = '0' if abs(x[0]) < 1e-3 else "%.3f"
5     txt = '\n%svalues=%"s' % (ind, xs)
6     for n in range(1, len(x)):
7         xs = '0' if abs(x[n]) < 1e-3 else "%.3f"
8         if len(txt) + 3 + len(xs) >= 80:
9             fid.write(txt + ';\n')
10            txt = ind + ' ' + xs
11        else:
12            txt += ' ; ' + xs
13 fid.write(txt + '\n')
```



## Test for code (R)

You can use the `string()` command to highlight a line of code.

```
1   write_list <- function(fid, x, level) {  
2     # Oh look, a comment!  
3     ind <- paste0(rep(" ", level), collapse = "")  
4     xs <- ifelse(abs(x[1]) < 1e-3, "0", "%.3f")  
5     txt <- sprintf("\n%values=%s", ind, xs)  
6  
7     for (n in 2:length(x)) { # wow!  
8       xs <- ifelse(abs(x[n]) < 1e-3, "0", "%.3f")  
9       if (nchar(txt) + 3 + nchar(xs) >= 80) {  
10         cat(txt, ";\\n", file = fid, sep = "")  
11         txt <- paste0(ind, " ", xs)  
12       } else {  
13         txt <- paste0(txt, "; ", xs)  
14       }  
15     }  
16  
17     cat(txt, "\\", file = fid, sep = "")  
18 }
```



# 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>