



UNIVERSIDADE DA CORUÑA

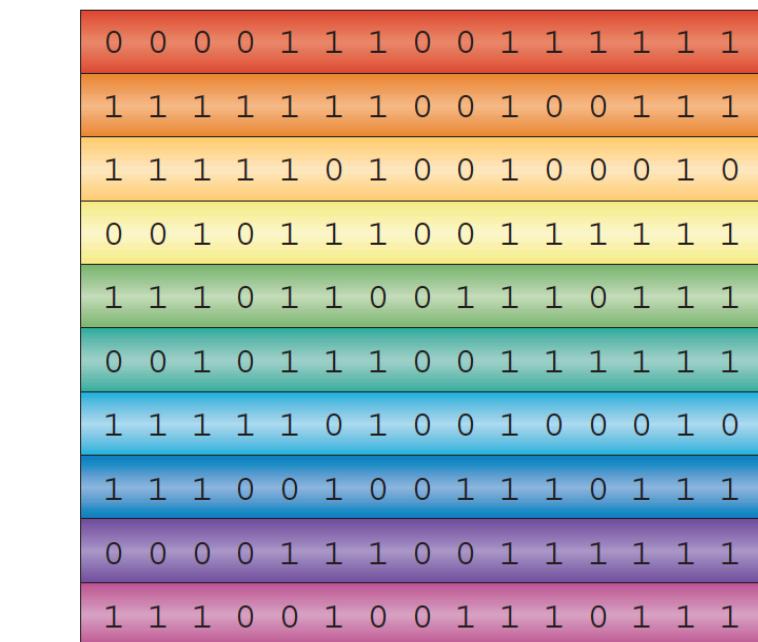
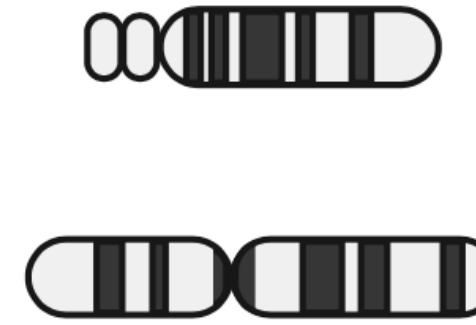


facultade de  
informática  
da coruña



Center for Research  
in Molecular Medicine and  
Chronic Diseases

# Comparación de la imputación genética en cohortes individuales *versus* imputación conjunta: Factores implicados y propuesta de actuación



Pedro Sánchez García

Directores Académicos: Raquel Cruz Guerrero y Silvia Díz de Almeida

Tutor Académico: Jorge González Domínguez

# ÍNDICE

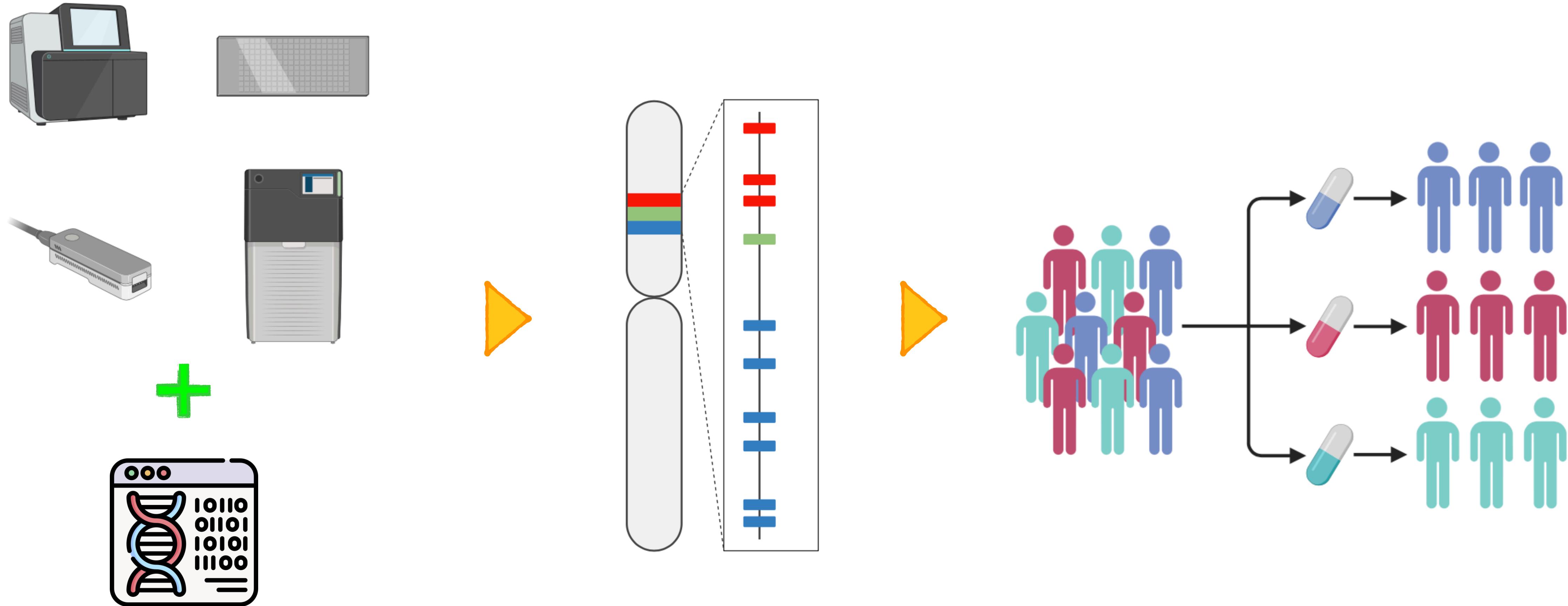
---

- 1. INTRODUCCIÓN.**
  
  - 2. OBJETIVOS.**
  
  - 3. MATERIALES Y MÉTODOS + RESULTADOS Y DISCUSIÓN.**
  
  - 4. CONCLUSIONES.**
-

# INTRODUCCIÓN

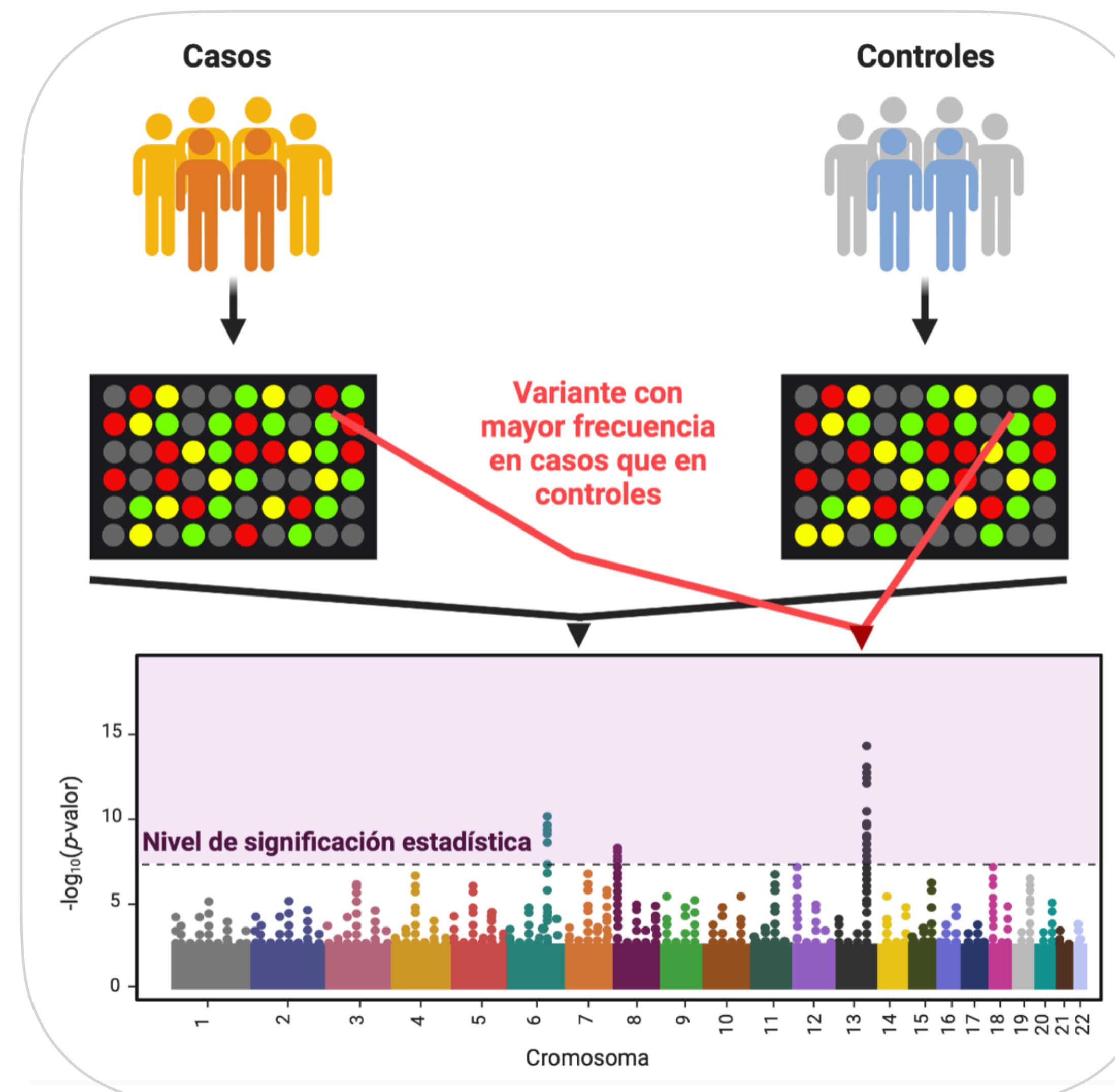
---

## / Contexto actual

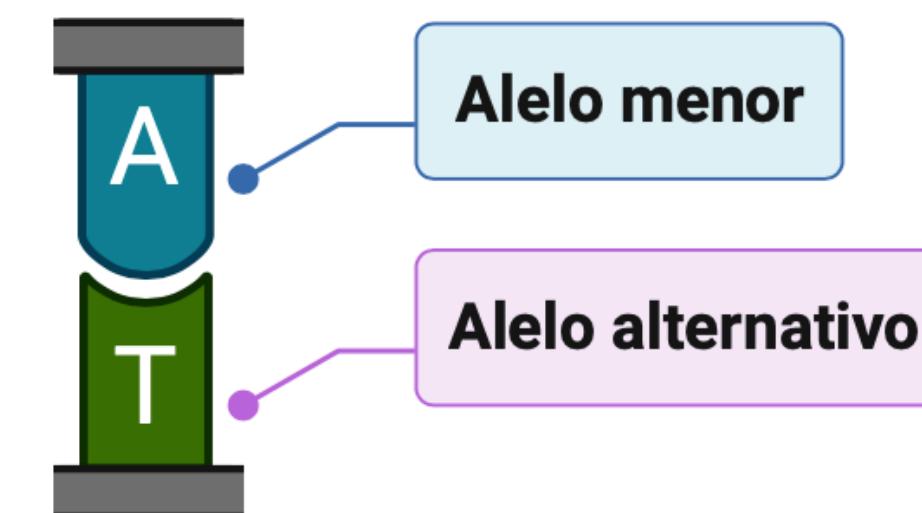


# INTRODUCCIÓN

## / Estudios de asociación del genoma completo (GWAS)



- ▶ Marcadores genéticos en **casos - controles**.
- ▶ Correlación genotipo - fenotipo: Asociaciones significativas.
- ▶ Variaciones de la frecuencia del alelo menor (**MAF**) y alelo alternativo en cromosoma para los grupos de pacientes:



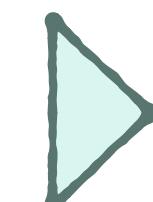
# INTRODUCCIÓN

## / Diseño experimental en estudios de asociación del genoma completo (GWAS)

### Control de calidad

✓ Eliminación de:

- Duplicados
- Muestras de baja calidad
- Errores de genotipado



### Imputación

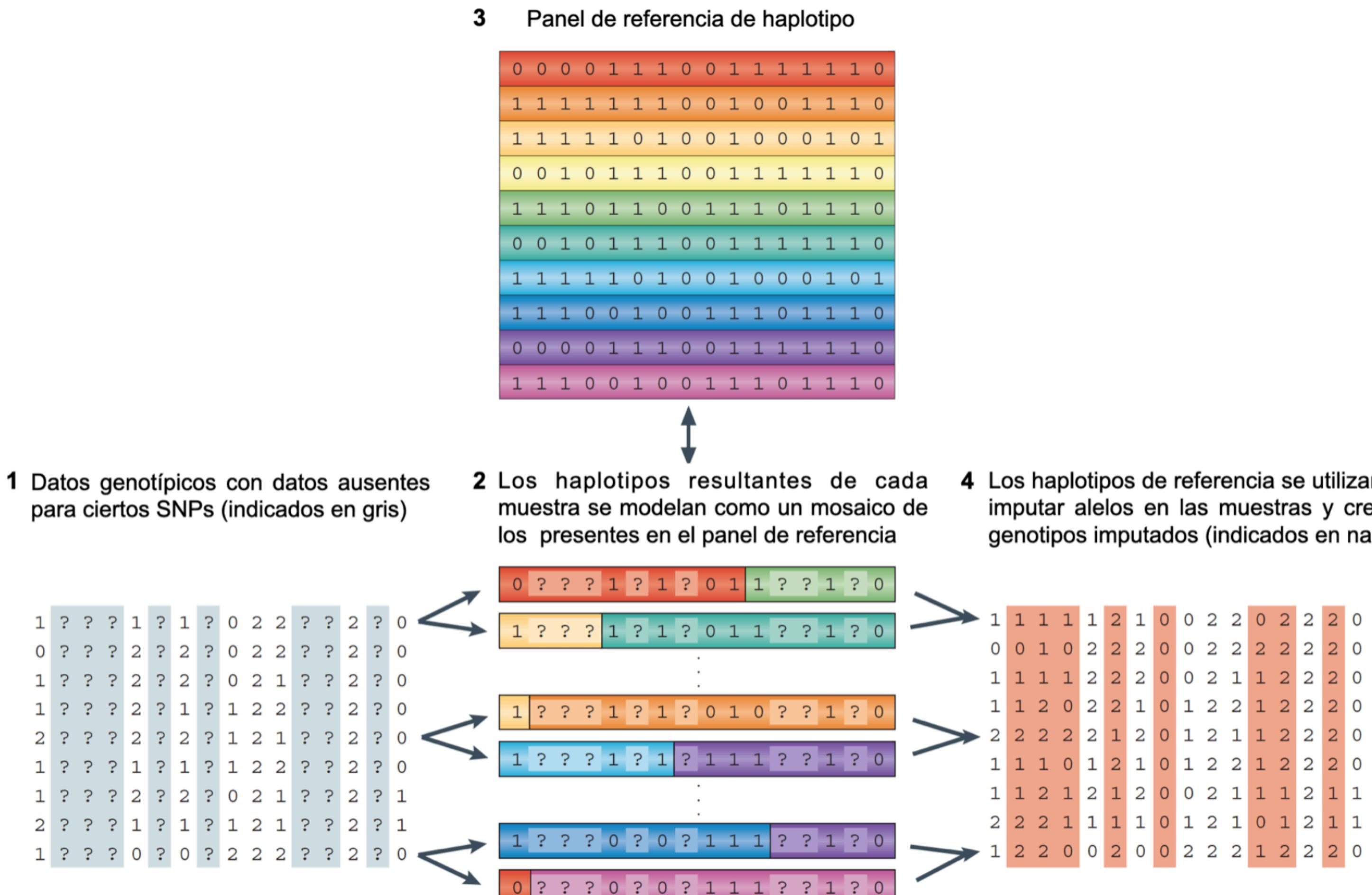
❖ Inferencia de marcadores genéticos ausentes en los *arrays* de genotipado

❖ Múltiples ventajas:

- Mayor resolución en asociaciones significativas
- Combinación de proyectos con diferentes *arrays*
- Replicación de resultados

# INTRODUCCIÓN

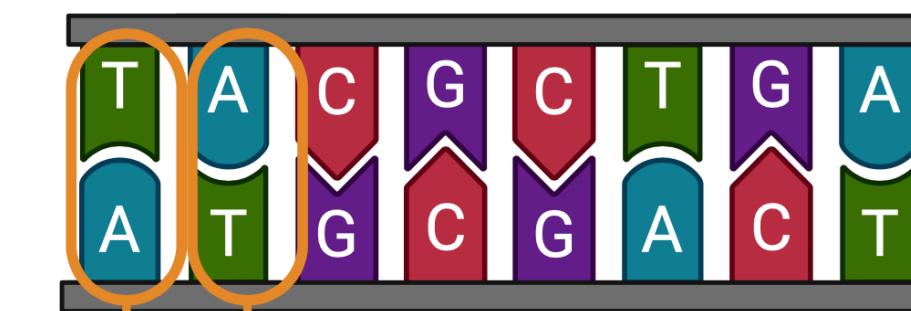
## / Aspectos principales de la imputación



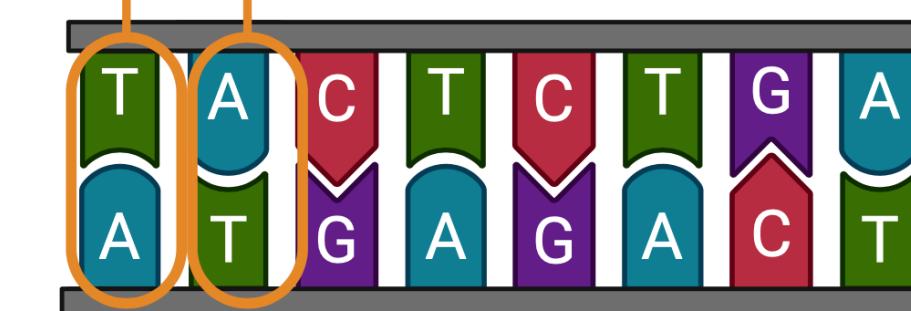
BIMBAM MaCH IMPUTE

Beagle Minimac

### Haplótipos



### Desequilibrio de ligamiento



BioData CATALYST  
TOPMed Imputation Server

- ✓ 50.000 individuos.
- ✓ 26 cohortes.

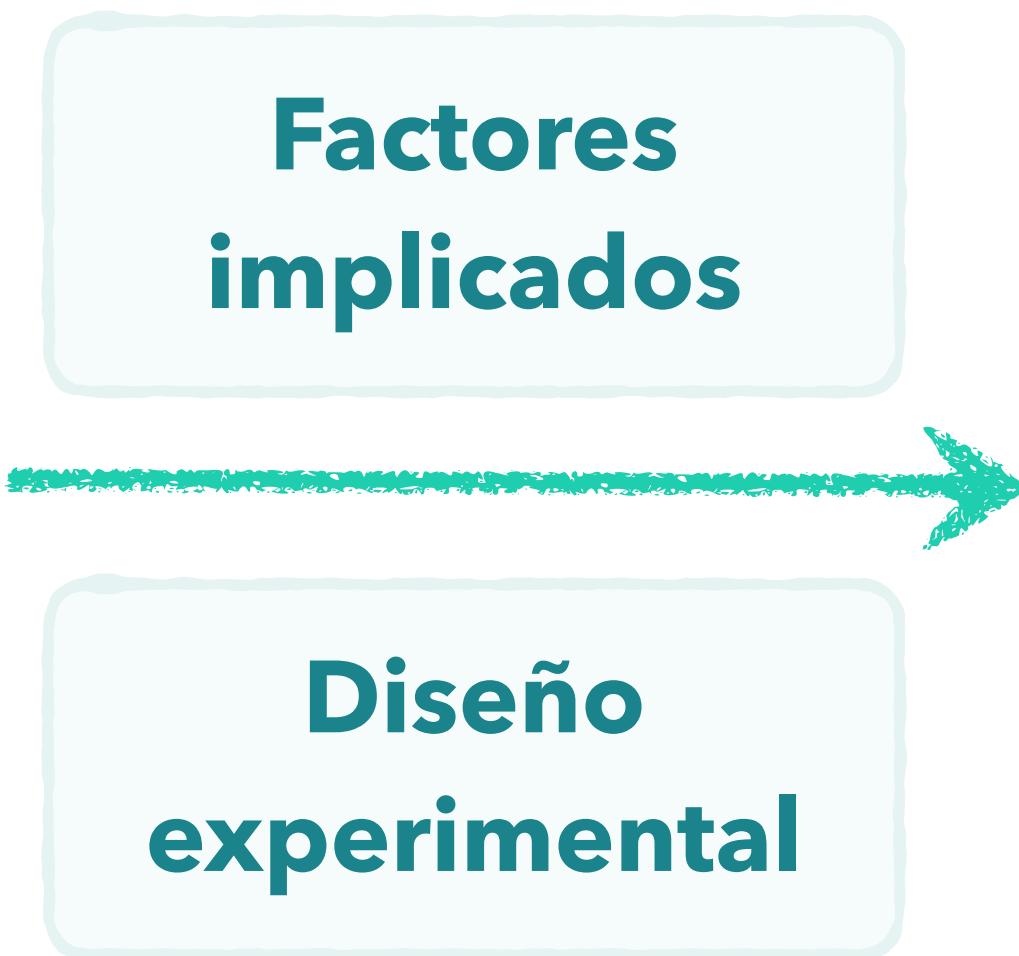
# INTRODUCCIÓN

## / Limitaciones de la imputación



# INTRODUCCIÓN

## / Esfuerzos en curso



**91% de los estudios  
realizados en individuos  
de ancestralidad europea**

**Ancestralidad mixta  
Complejidad en estructura genética**

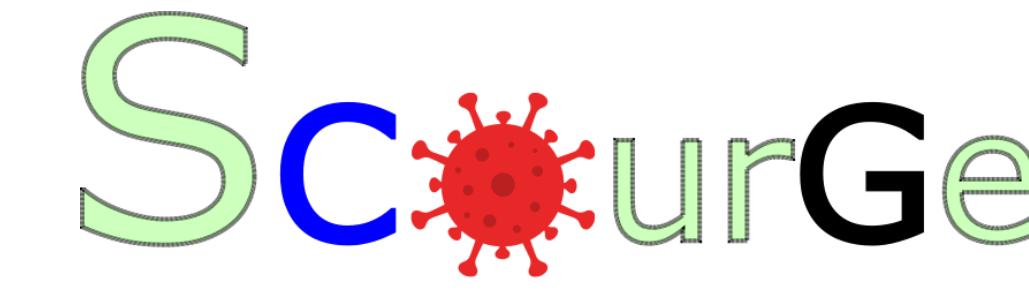
# OBJETIVOS

---

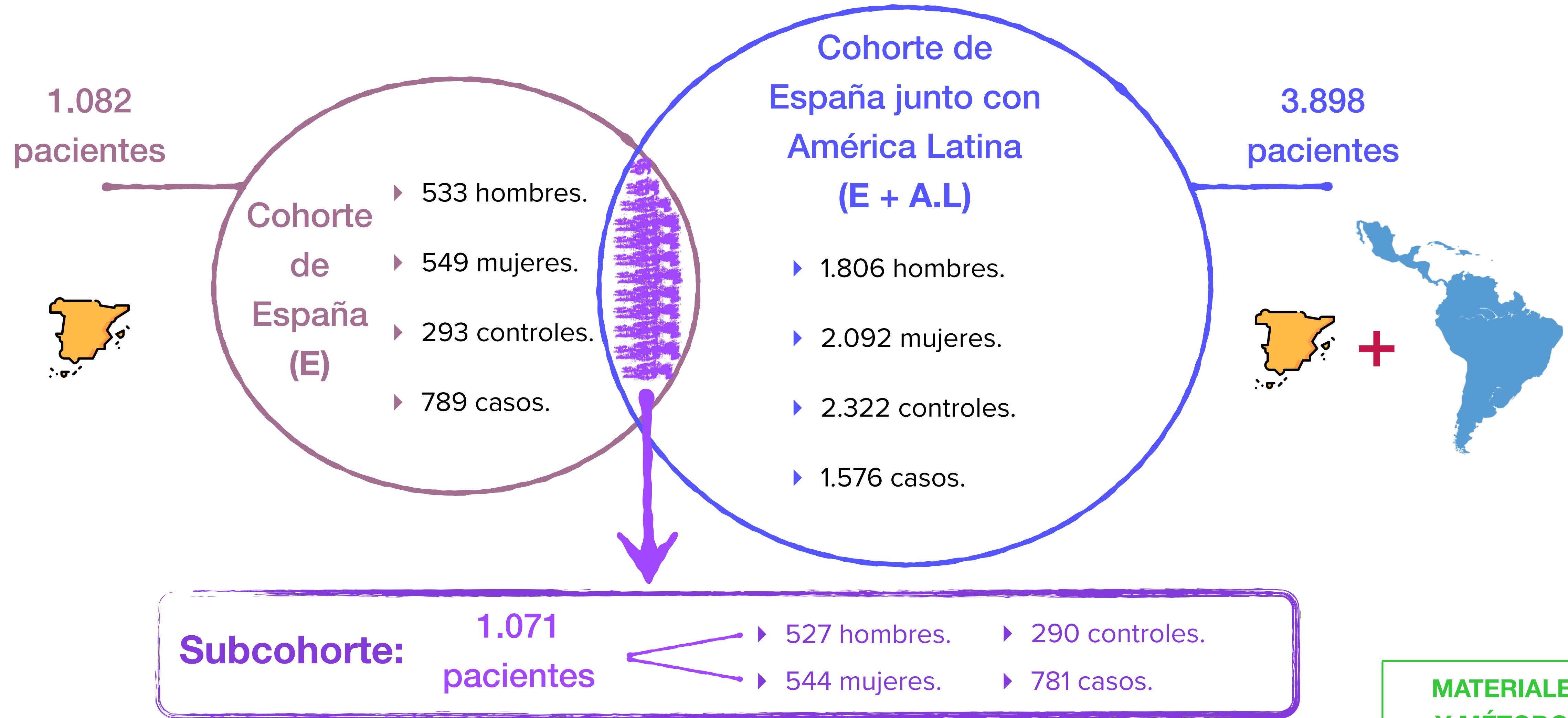
**Caracterizar las diferencias en la imputación realizada en una subcohorte de pacientes latinos en dos situaciones: cuando se imputa de forma individual o en conjunto con el total de la cohorte.**

- Realizar un análisis exploratorio de factores que potencialmente influyen en la diferencia en estima de frecuencias alélicas: distribución y densidad de marcadores en el *array* de genotipado.
- Determinar un criterio de filtrado o control de calidad posterior a la imputación para obtener unos resultados fiables en un estudio típico.

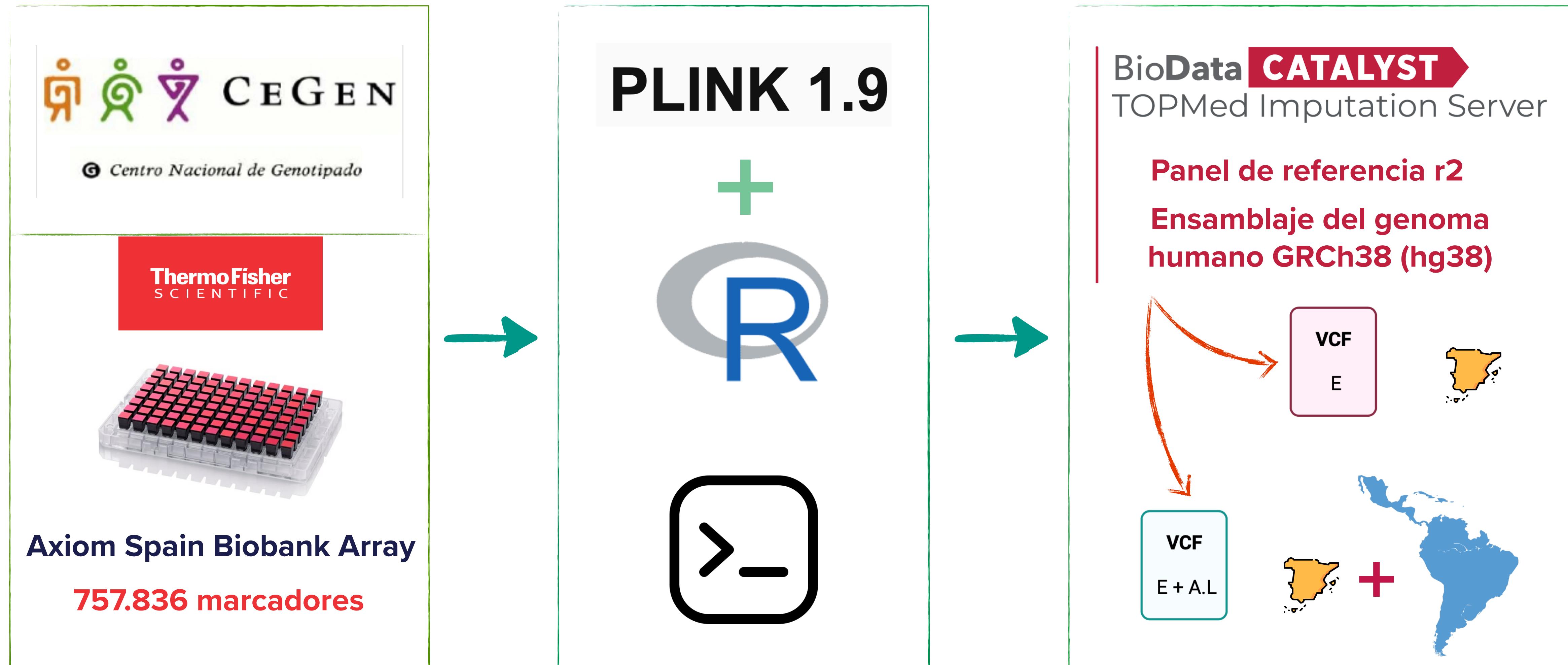
# I Cohortes de estudio



Spanish COalition to Unlock Research  
on host GEnetics on COVID-19

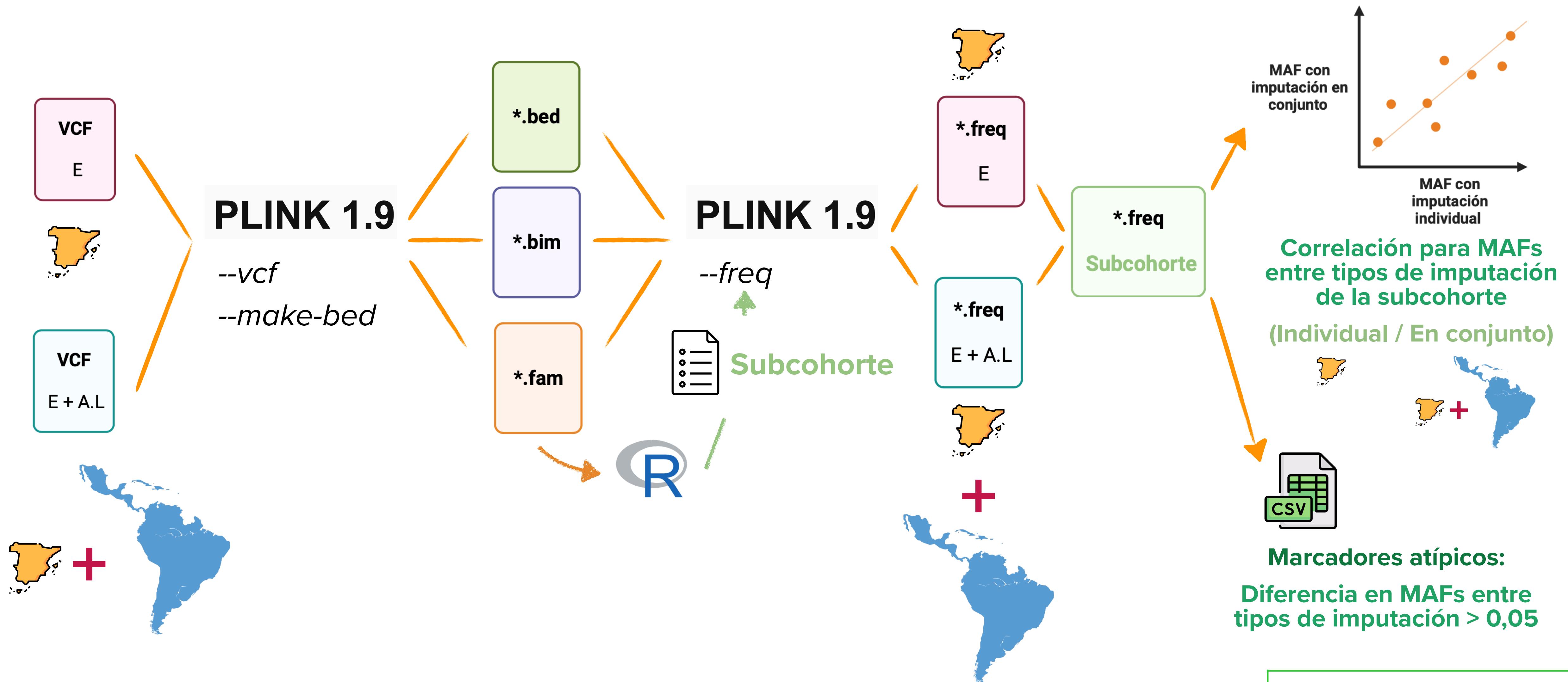


# Genotipado, control de calidad e imputación

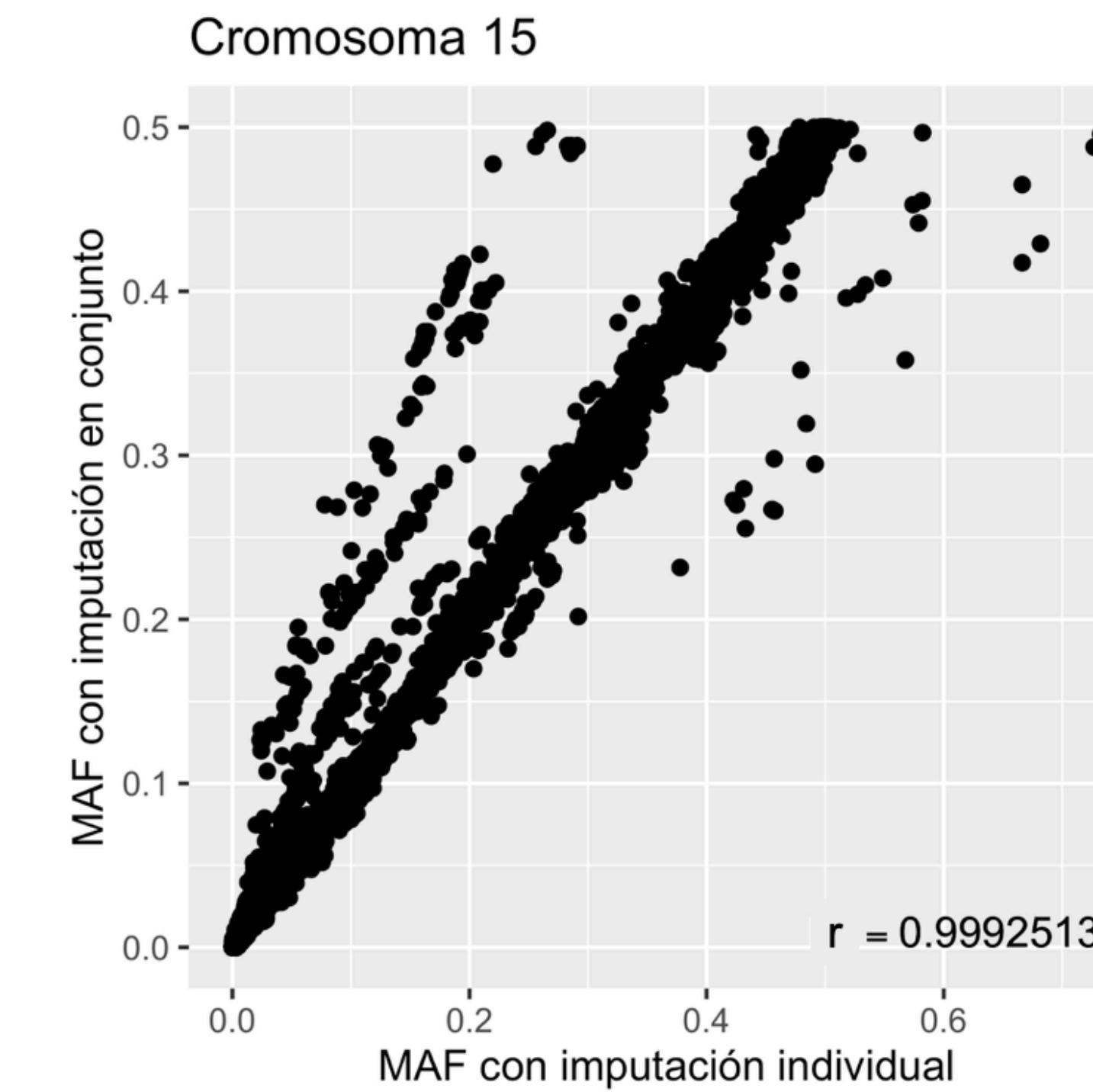
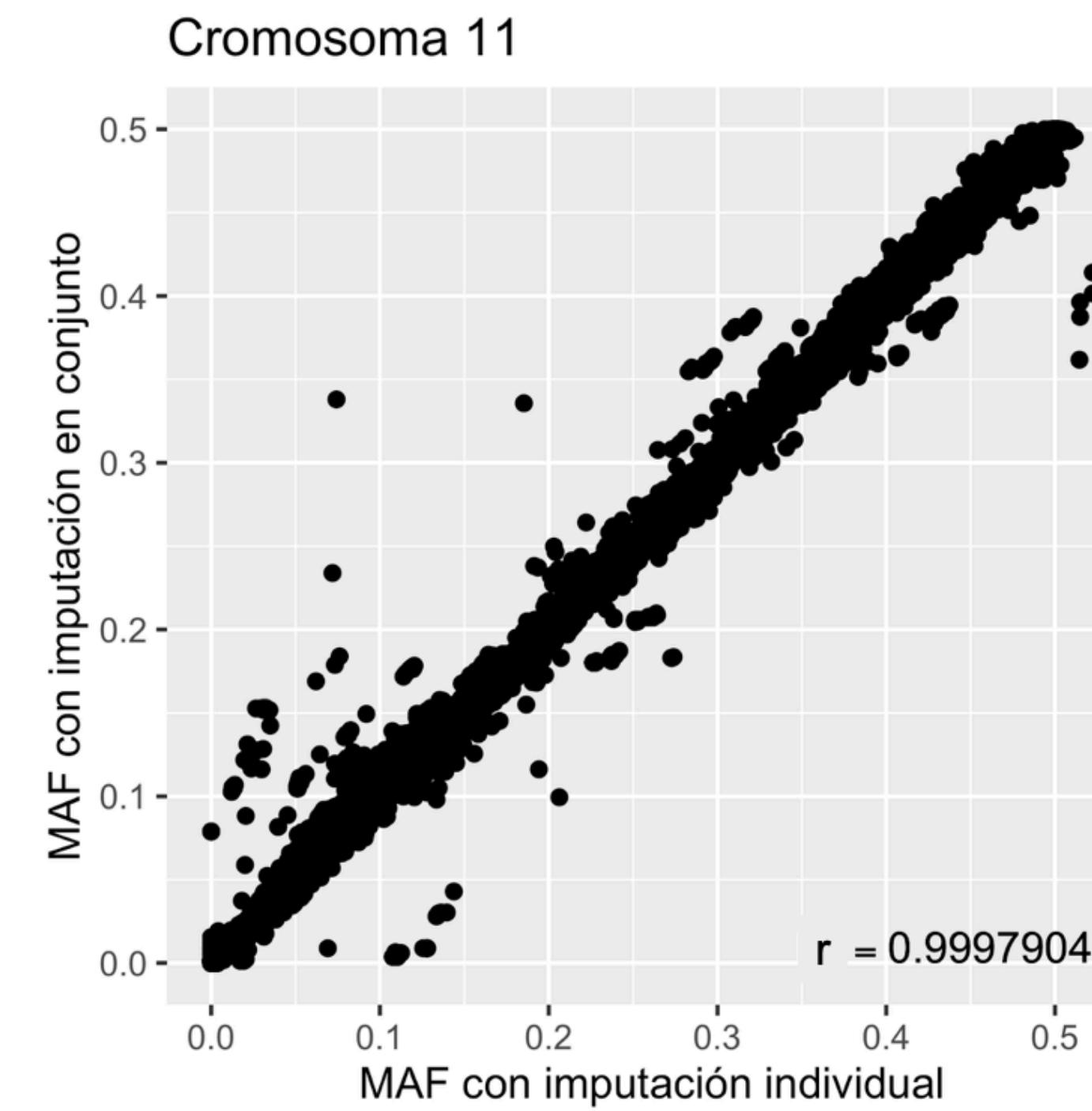


MATERIALES Y MÉTODOS

# Comparativa de las MAFs entre tipos de imputación



# | Comparativa de las MAFs entre tipos de imputación



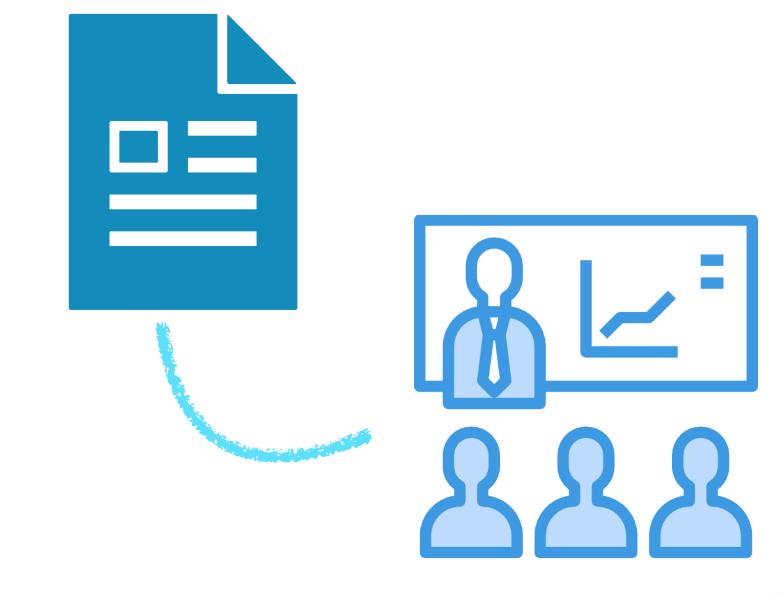
- ❖ Fuerte dependencia lineal de MAFs entre tipos de imputación de la subcohorte ( $r \approx 1$ ).
- ❖ Numerosos marcadores con importantes diferencias en MAFs según tipo de imputación.

# Comparativa de las MAFs entre tipos de imputación

Cromosoma	Marcadores totales	Marcadores atípicos	r
1	8.553.267	219	0,9997864
2	9.223.282	140	0,9999255
3	7.655.401	2	0,9999565
4	7.561.380	7	0,9999548
5	<b>7.005.543</b>	-----	<b>0,9999601</b>
6	6.668.396	4	0,9999552
7	6.275.350	16	0,9999506
8	6.000.476	1	0,9999560
9	4.704.138	162	0,9997886
10	5.157.212	6	0,9999370
11	<b>5.270.952</b>	<b>357</b>	<b>0,9997904</b>
12	<b>5.133.523</b>	-----	<b>0,9999519</b>
13	3.862.684	6	0,9999565
14	3.431.884	1	0,9999380
15	<b>3.147.065</b>	<b>302</b>	<b>0,9992513</b>
16	3.480.859	11	0,9999121
17	3.073.840	10	0,9999306
18	3.066.725	11	0,9999518
19	2.407.721	1	0,9999429
20	2.419.620	1	0,9999550
21	1.430.516	82	0,9998406
22	1.468.096	69	0,9997846
X	3.982.189	93	0,9998425

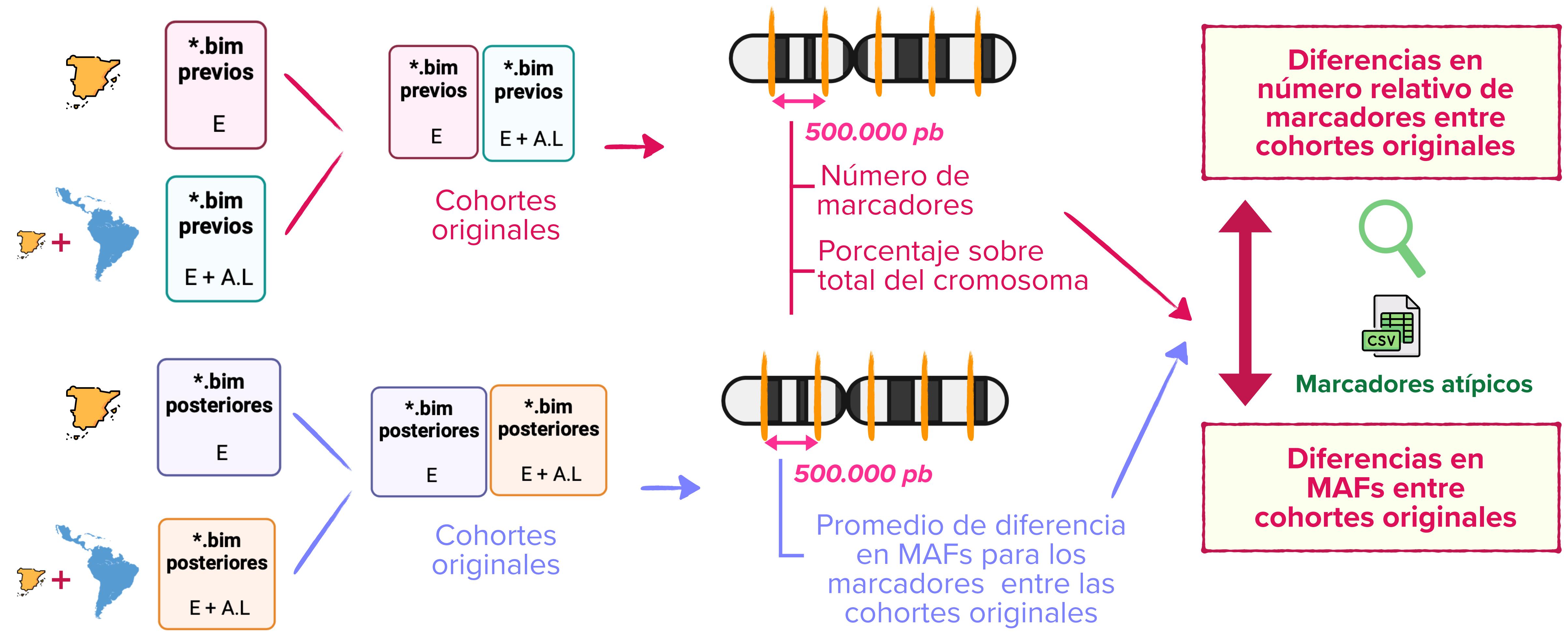
Ausencia de marcadores atípicos

Mayor número de marcadores atípicos

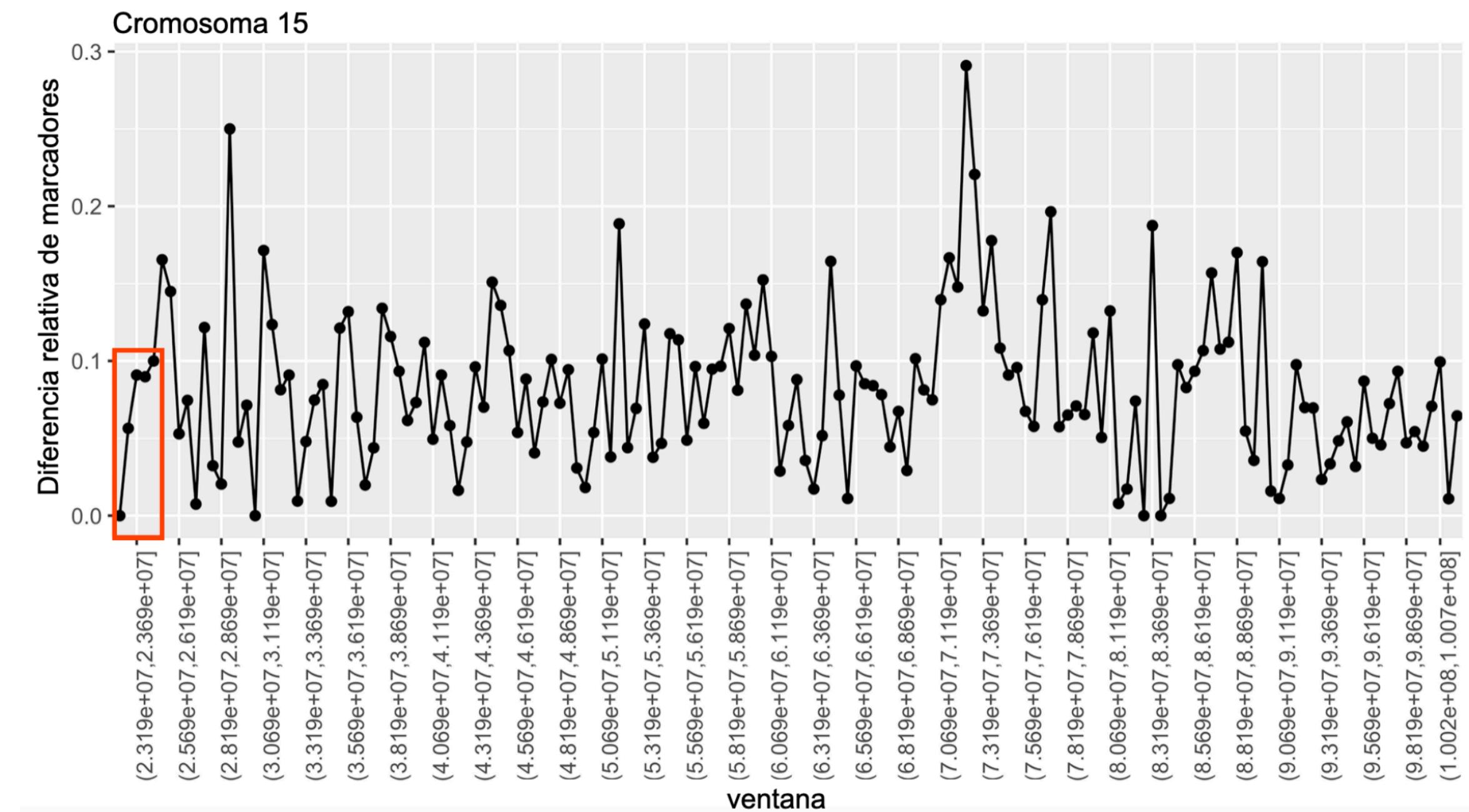
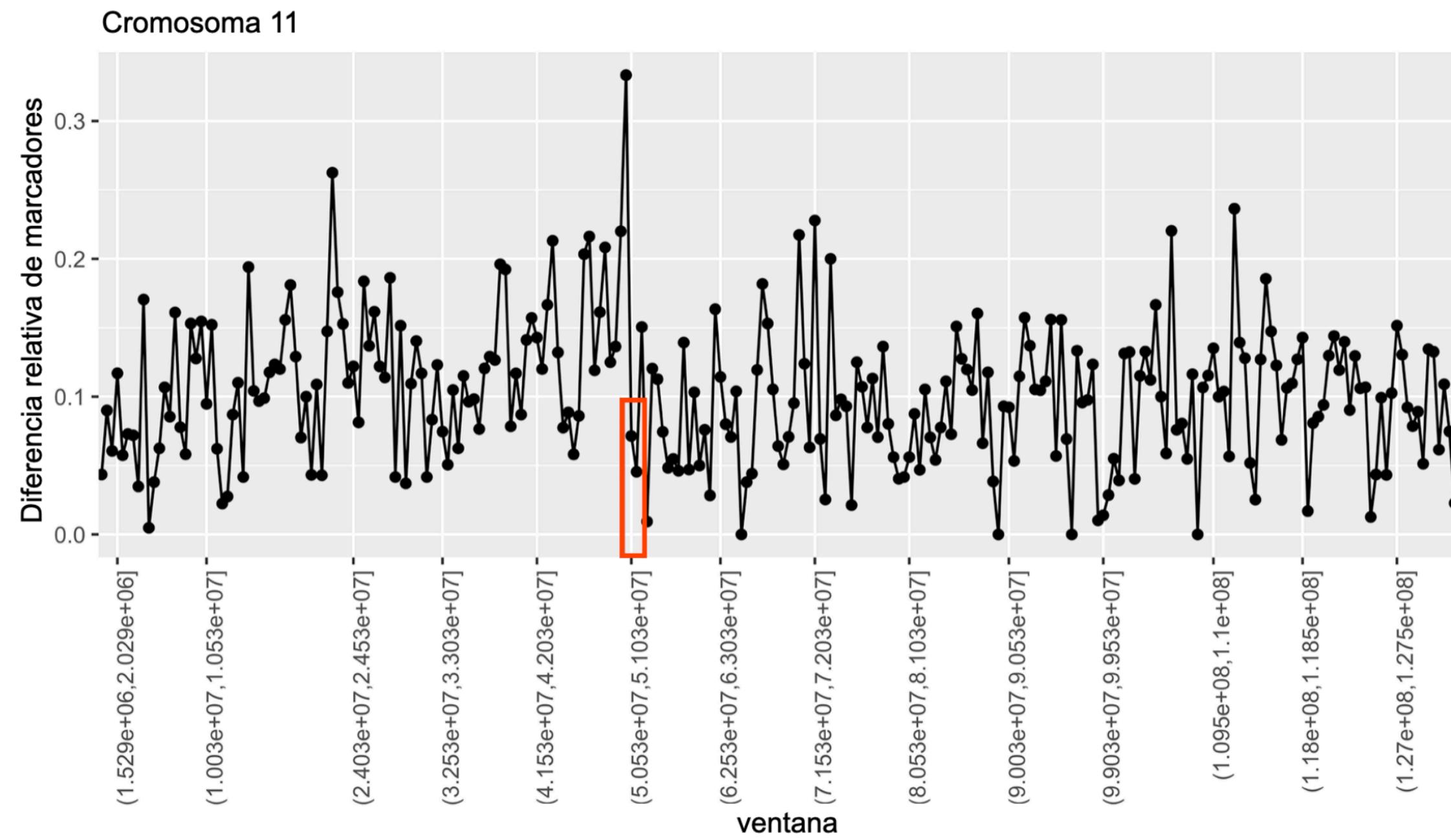


RESULTADOS Y DISCUSIÓN

# | Análisis de la influencia en la estima de MAFs por densidad y distribución de marcadores

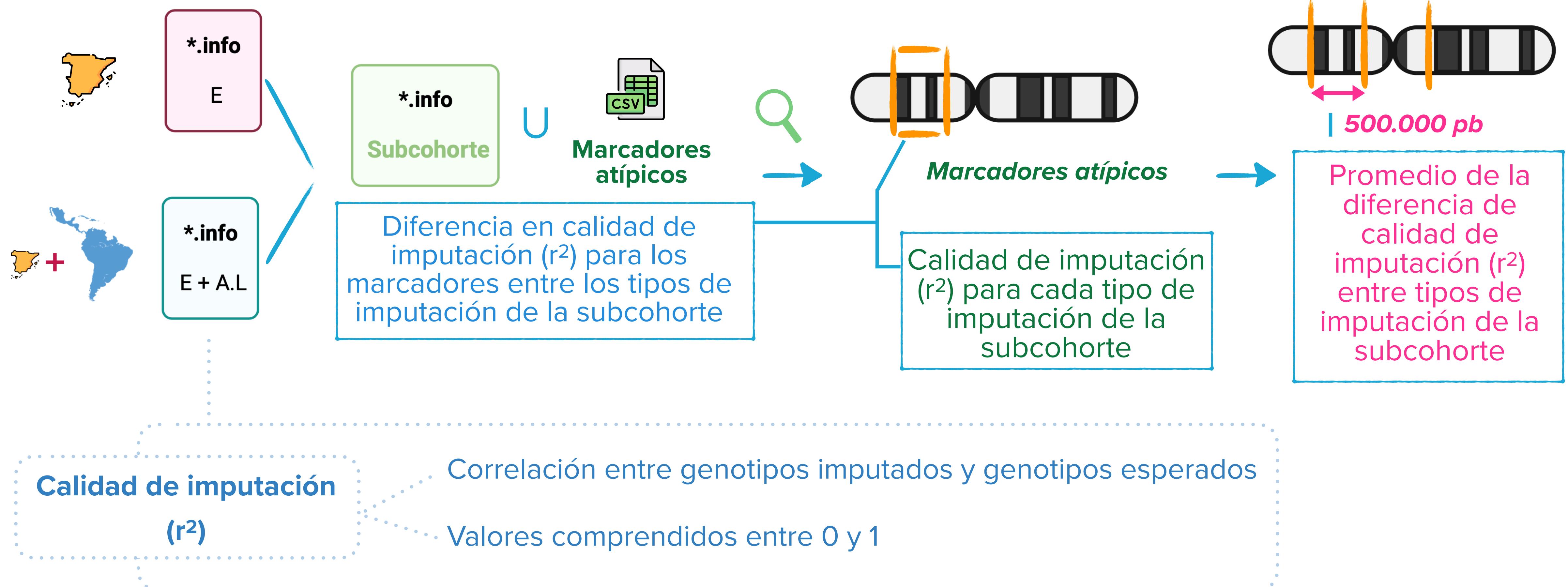


# | Análisis de la influencia en la estima de MAFs por densidad y distribución de marcadores



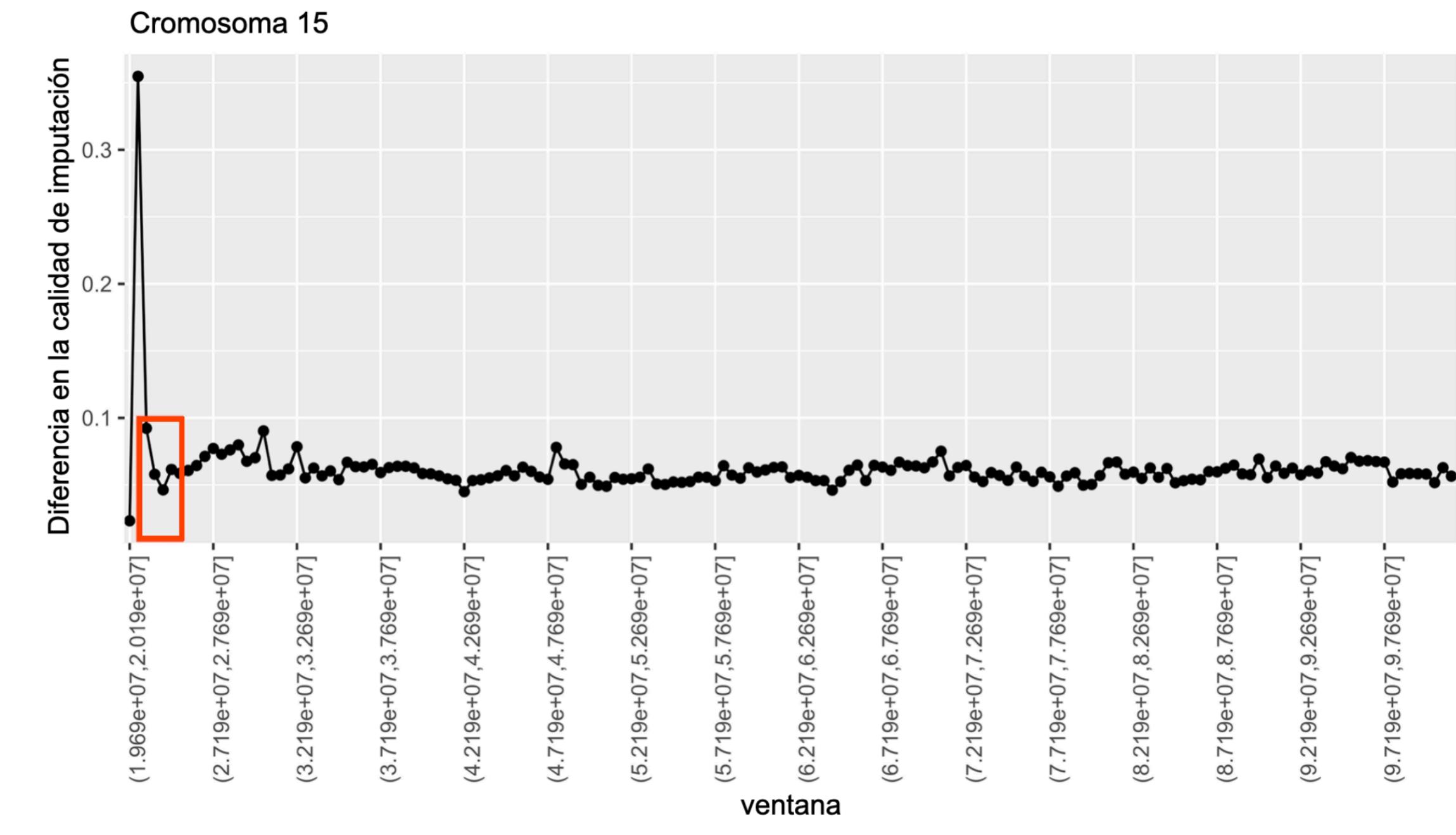
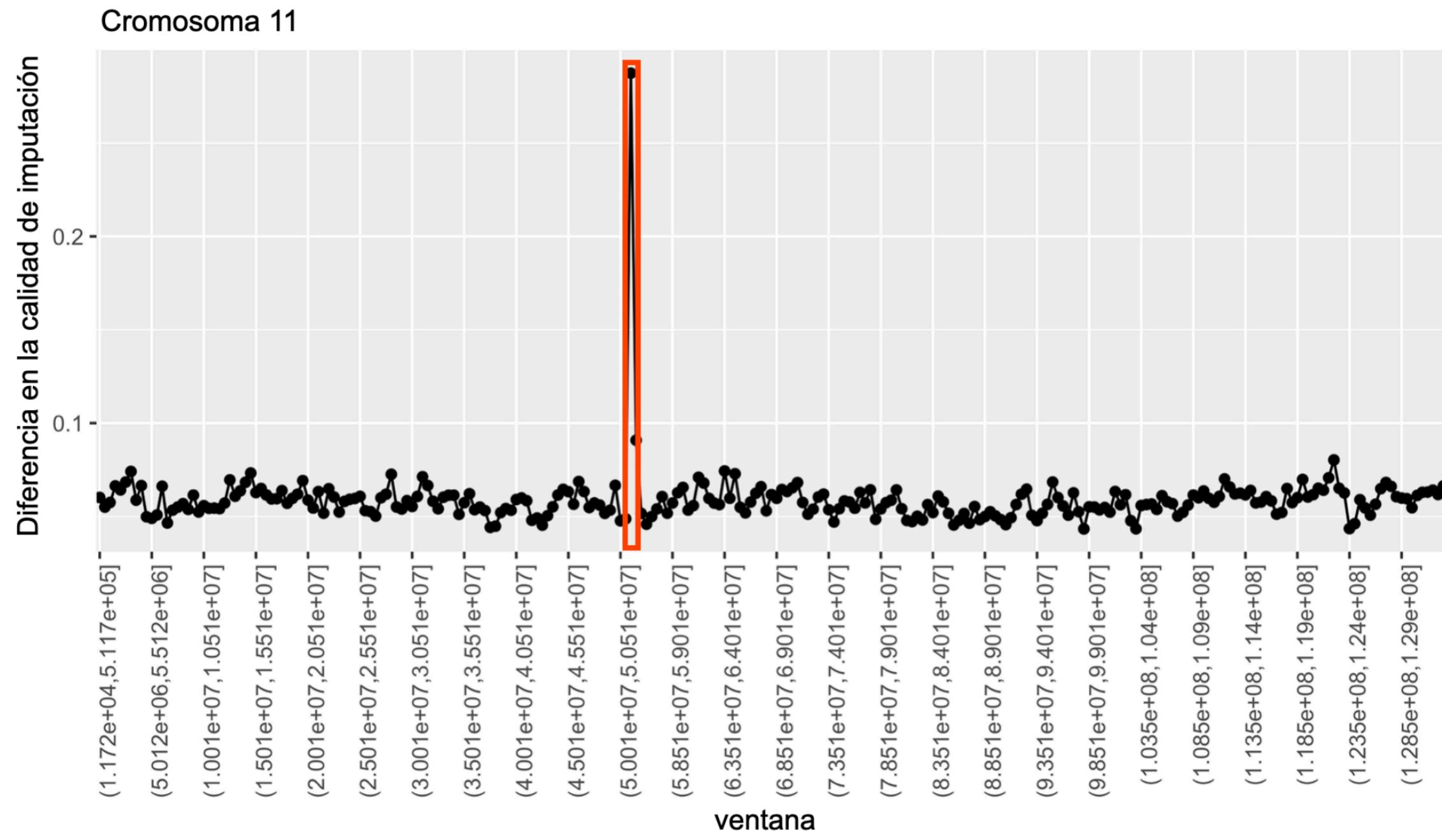
- ❖ Notables variaciones de diferencias en número relativo de marcadores entre cohortes originales.
- ❖ Ausencia de valores elevados en las **ventanas** con los **marcadores atípicos en términos de MAFs**.

# | Relevancia de la calidad de imputación en la estimación de MAFs



# | Relevancia de la calidad de imputación en la estimación de MAFs

/ Promedio de la diferencia en calidad de imputación ( $r^2$ ) entre tipos de imputación de subcohorte

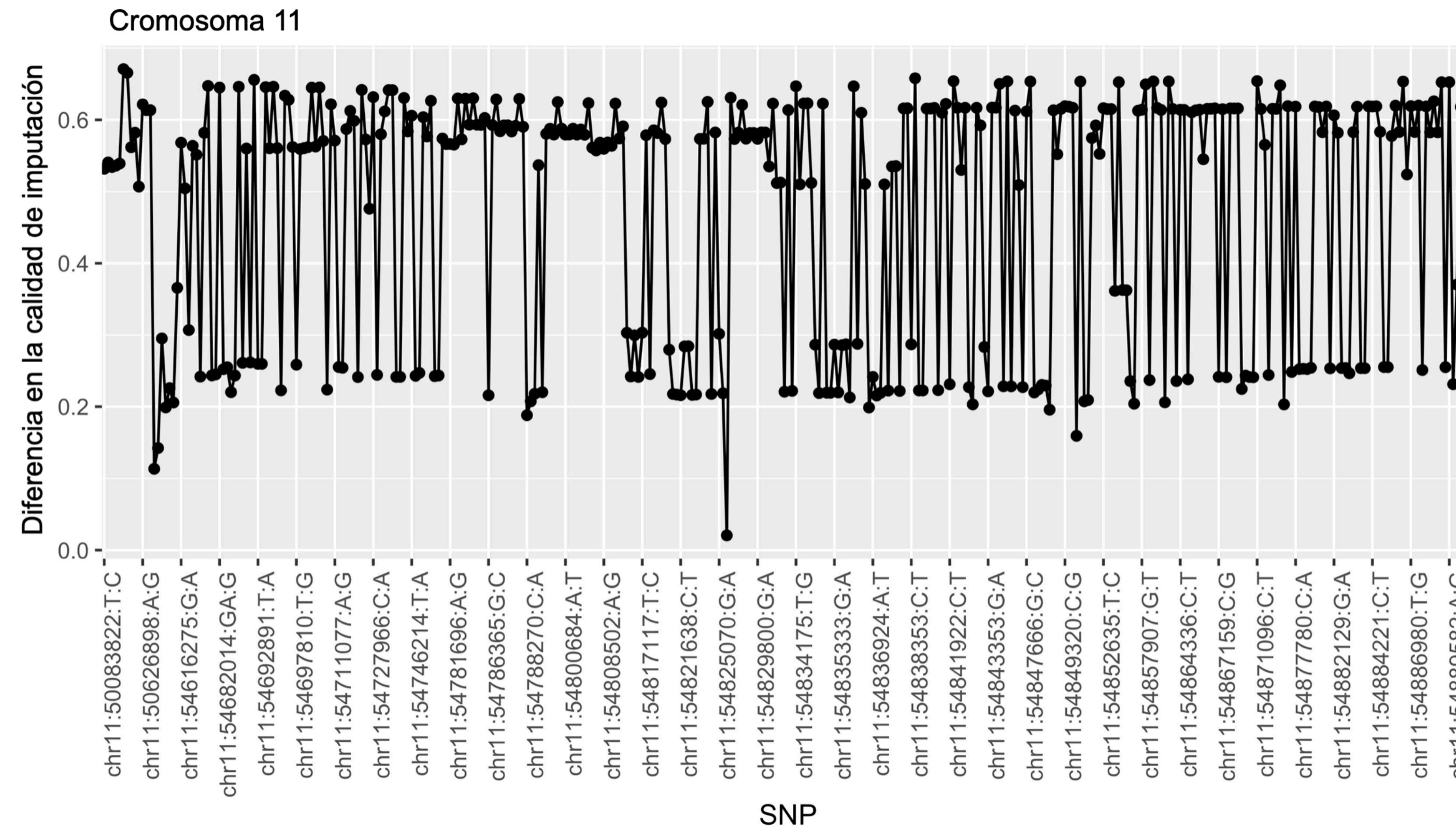


Máxima diferencia en calidad de imputación ( $r^2$ ) en las ventanas con los marcadores atípicos

RESULTADOS Y DISCUSIÓN

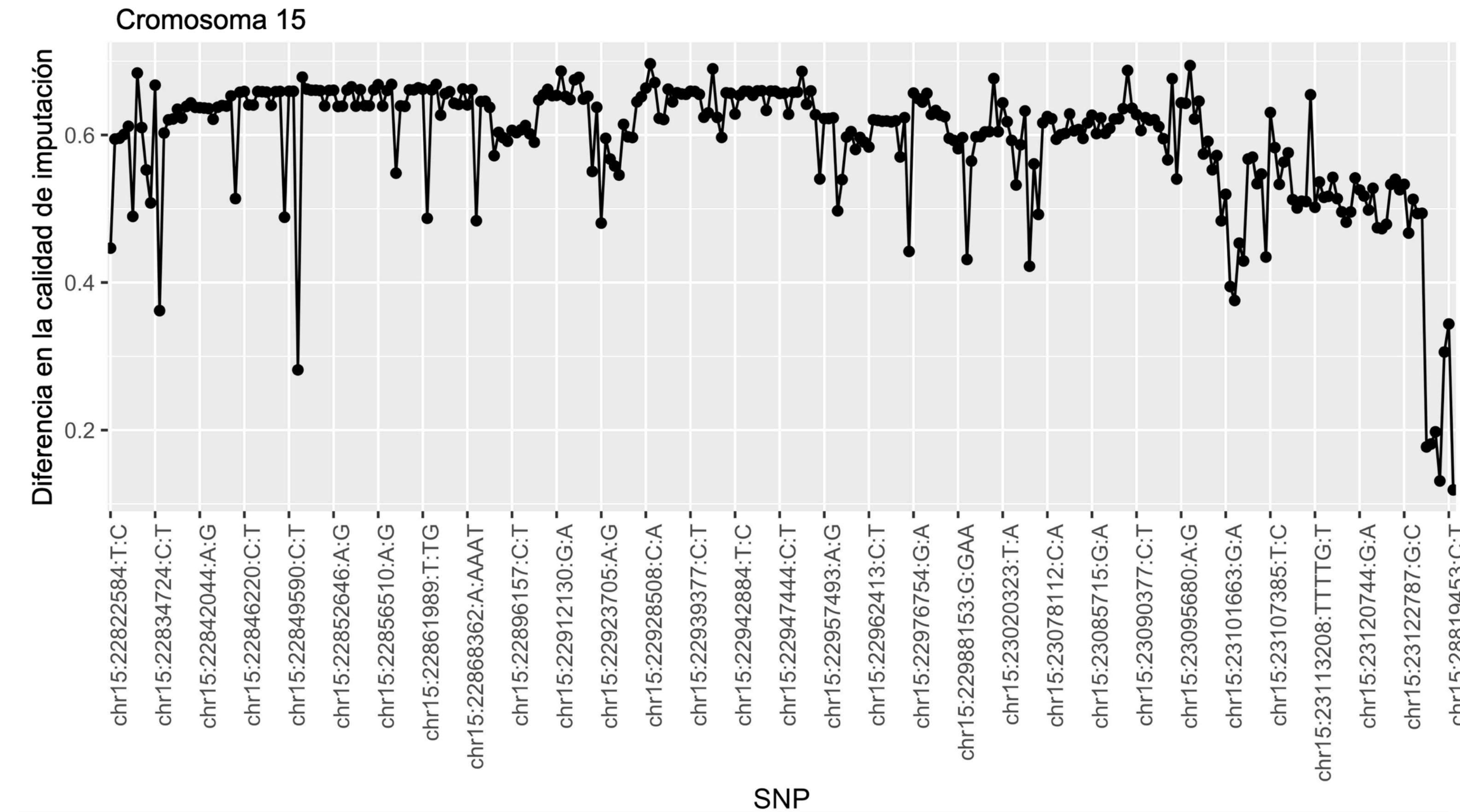
# | Relevancia de la calidad de imputación en la estimación de MAFs

/ Diferencia en calidad de imputación ( $r^2$ ) para los marcadores atípicos



# | Relevancia de la calidad de imputación en la estimación de MAFs

/ Diferencia en calidad de imputación ( $r^2$ ) para los marcadores atípicos



RESULTADOS Y DISCUSIÓN

# | Relevancia de la calidad de imputación en la estimación de MAFs

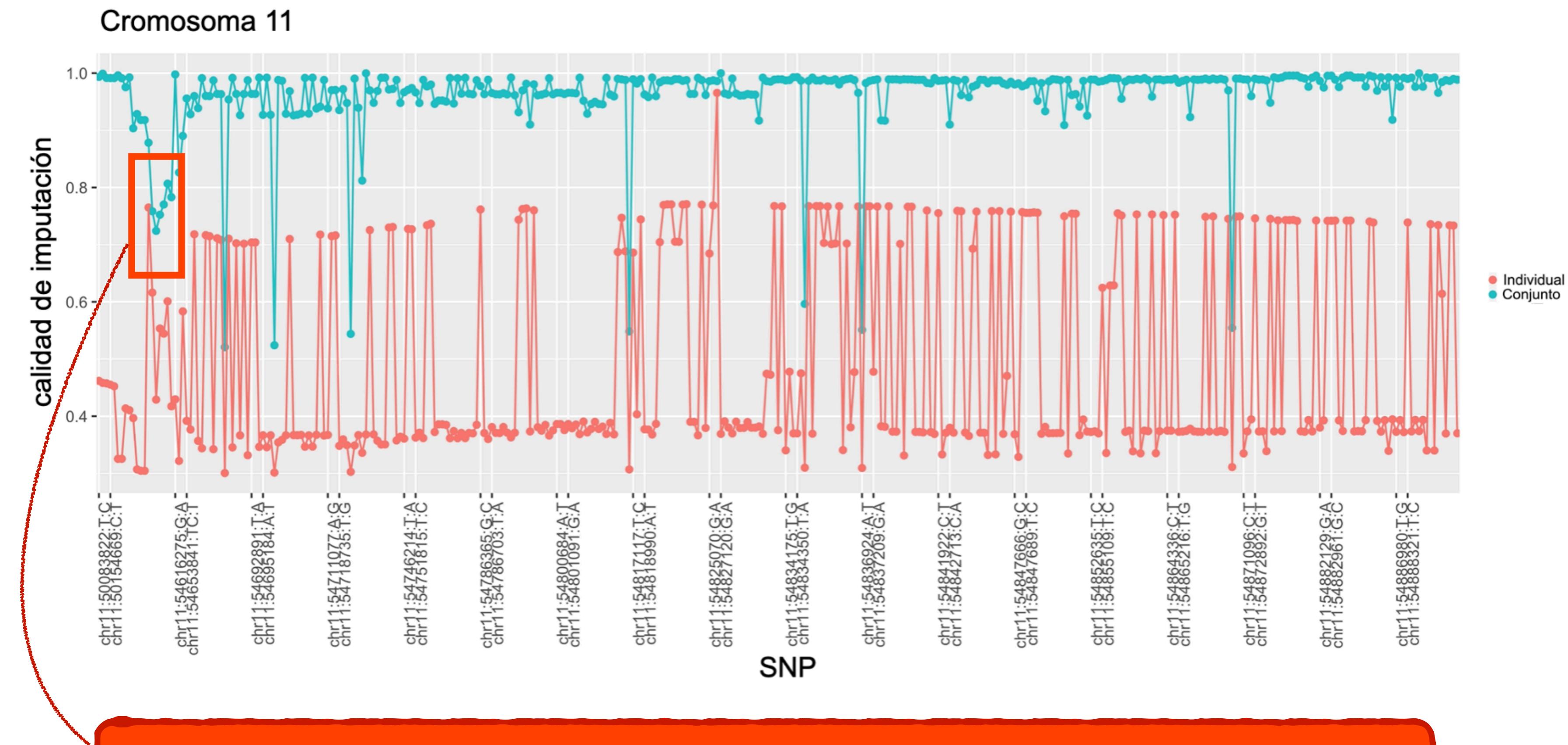
## / Calidad de imputación ( $r^2$ ) en los marcadores atípicos según tipo de imputación de la subcohorte



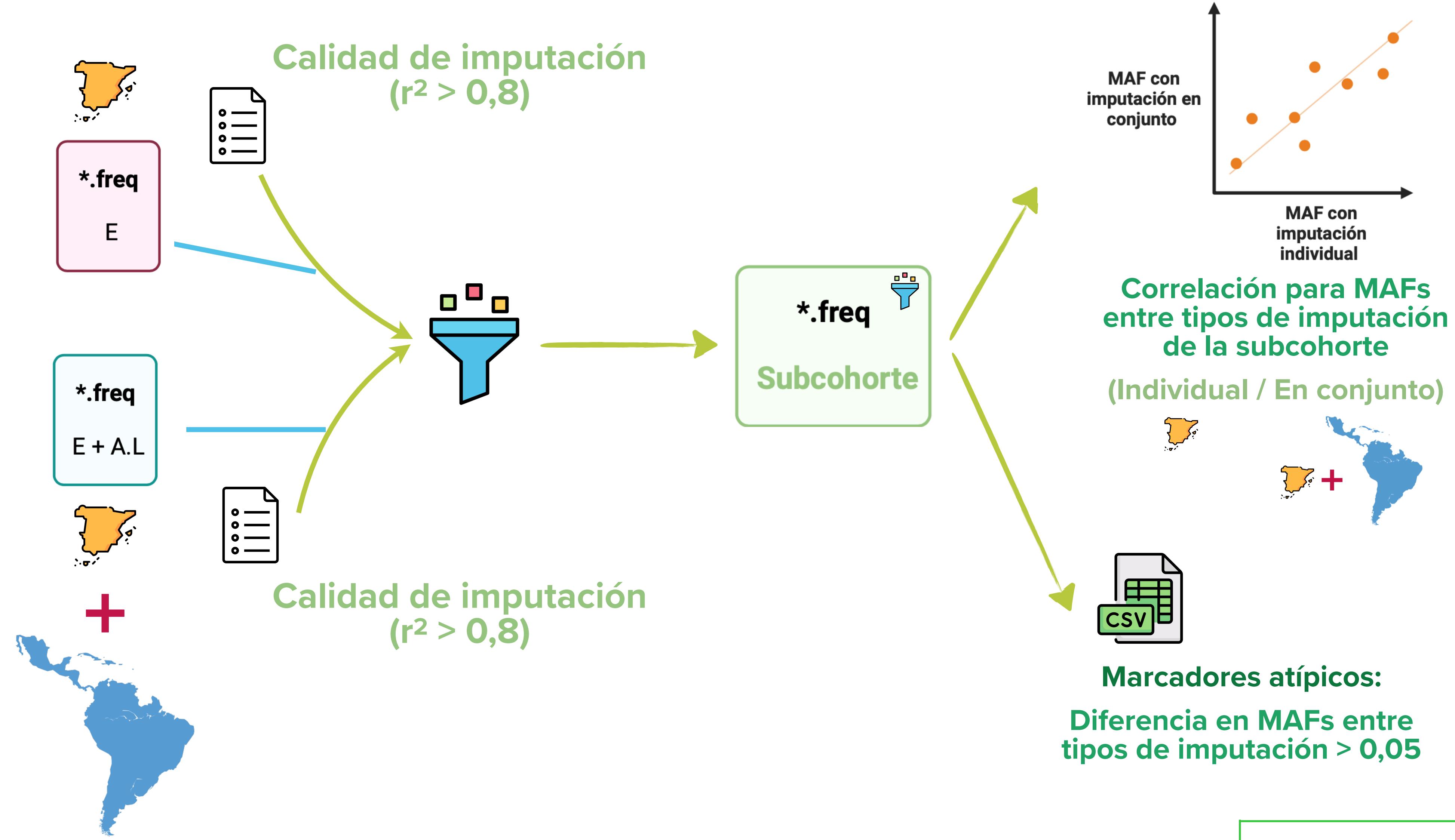
# RESULTADOS Y DISCUSIÓN

# I Relevancia de la calidad de imputación en la estimación de MAFs

/ Calidad de imputación ( $r^2$ ) en los marcadores atípicos según tipo de imputación de la subcohorte

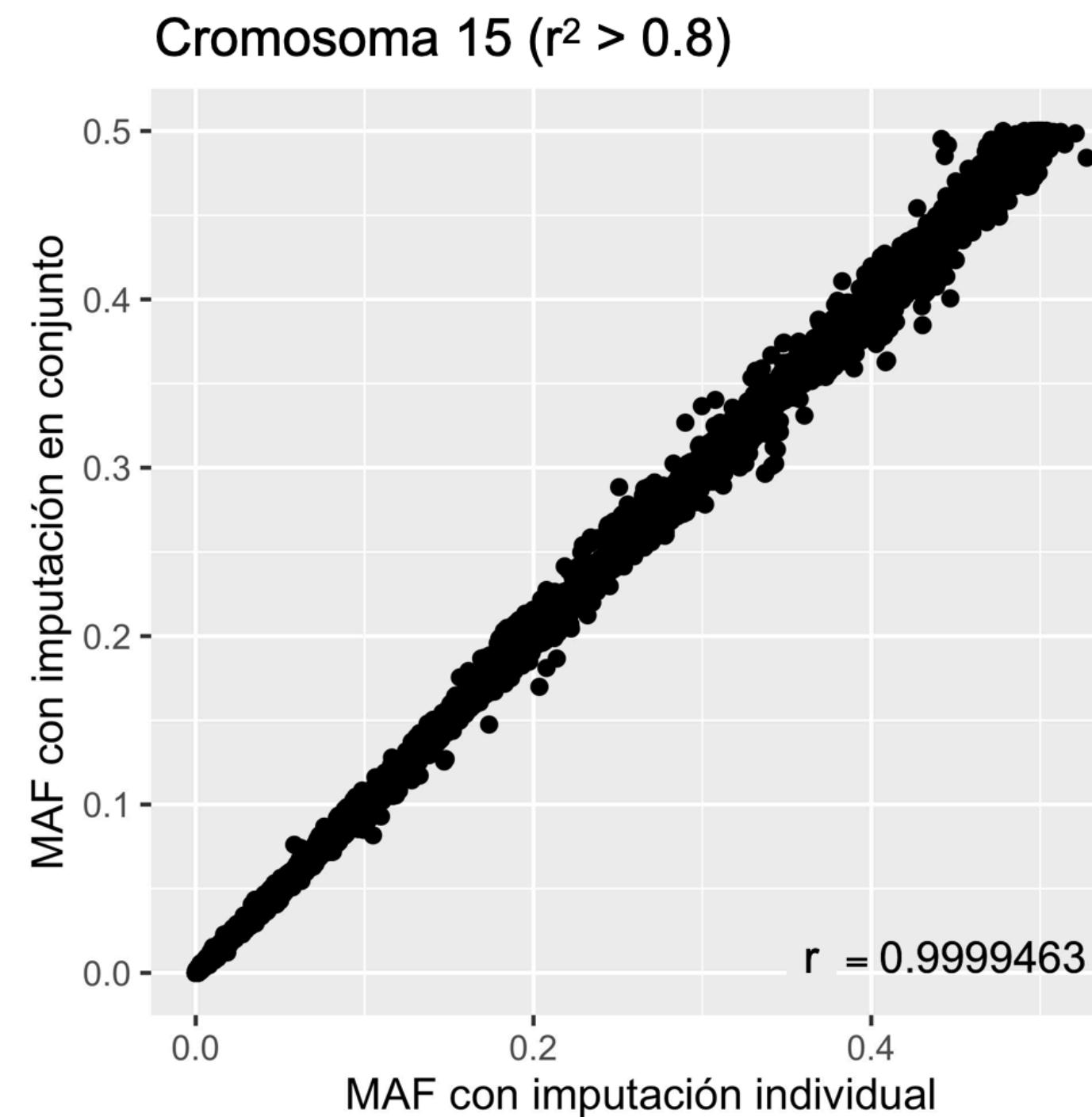
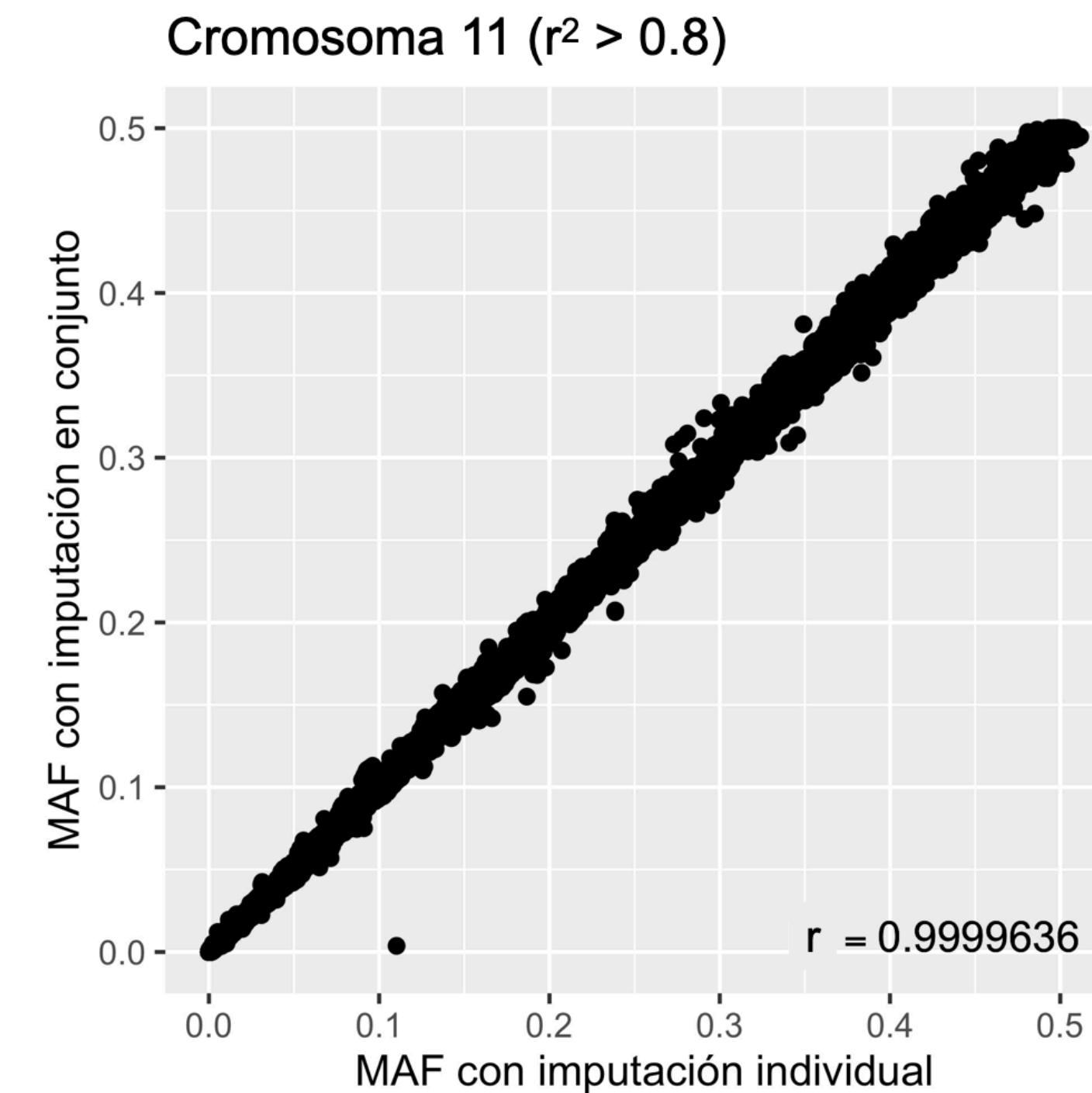


# | Filtrado de la calidad de imputación



# | Filtrado de la calidad de imputación

## / Comparativa de las MAFs entre tipos de imputación



- ◆ Mejora en dependencia lineal de MAFs entre tipos de imputación de la subcohorte ( $r \approx 1$ ).
- ◆ Drástico descenso de marcadores con importantes diferencias en MAFs según tipo de imputación.

# | Filtrado de la calidad de imputación

## / Comparativa de las MAFs entre tipos de imputación

**Marcadores filtrados (< 50%)**

Poder estadístico



**Balance en calidad de imputación**

↑ Umbral

Solidez en posteriores análisis

**RESULTADOS Y DISCUSIÓN**

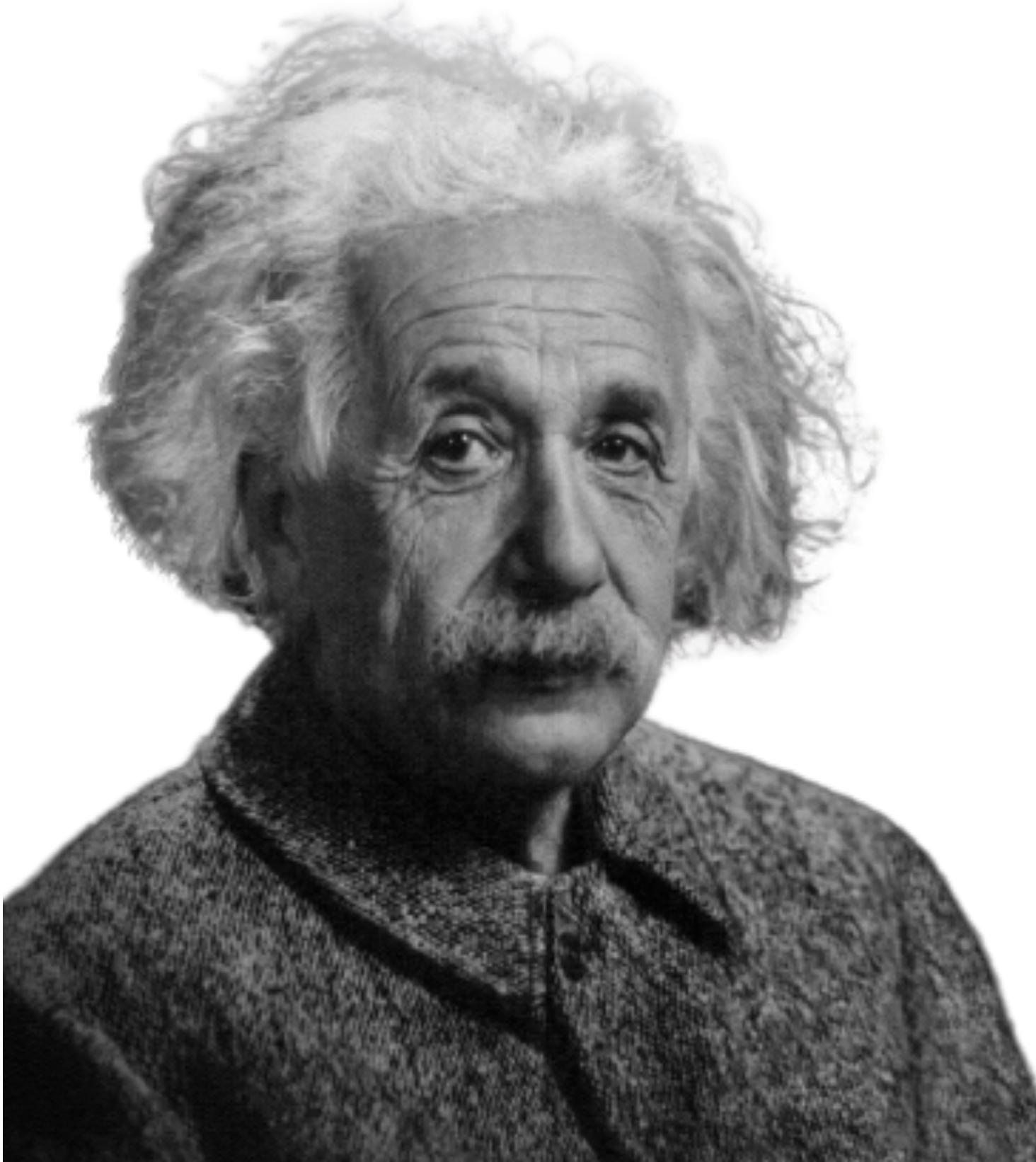
Cromosoma	Marcadores totales	Marcadores filtrados	Marcadores atípicos	r
1	8.553.267	6.009.109	-	0,9999602
2	9.223.282	6.529.383	-	0,9999525
3	7.655.401	5.466.344	-	0,9999662
4	7.561.380	5.413.192	-	0,9999651
6	6.668.396	4.840.996	-	0,9999636
7	6.275.350	4.457.000	-	0,9999677
8	6.000.476	4.284.320	-	0,9999639
9	4.704.138	3.316.225	-	0,9999550
10	5.157.212	3.755.212	-	0,9999611
11	5.270.952	3.743.241	1	0,9999636
13	3.862.684	2.745.075	-	0,9999646
14	3.431.884	2.428.213	-	0,9999576
15	3.147.065	2.204.444	-	0,9999463
16	3.480.859	2.388.588	3	0,9999437
17	3.073.840	2.096.336	-	0,9999641
18	3.066.725	2.156.883	-	0,9999632
19	2.407.721	1.651.122	-	0,9999558
20	2.419.620	1.667.101	-	0,9999642
21	1.430.516	978.949	-	0,9999618
22	1.468.096	987.398	1	0,9999474
X	3.982.189	2.439.908	6	0,9999444

# CONCLUSIONES

---

- I. Se producen notables diferencias en MAFs para numerosos marcadores según el tipo de imputación en la subcohorte de individuos latinos analizada.
- II. La calidad de imputación representa el factor causal de las diferencias en MAFs observadas.
- III. Fijar un filtro estricto en la calidad de imputación puede definir un criterio eficiente en aquellos estudios internacionales que impliquen la imputación de poblaciones de latinos.

# MUCHAS GRACIAS



**“Lo importante es no dejar de  
hacerse preguntas”**

**Albert Einstein**



UNIVERSIDADE DA CORUÑA

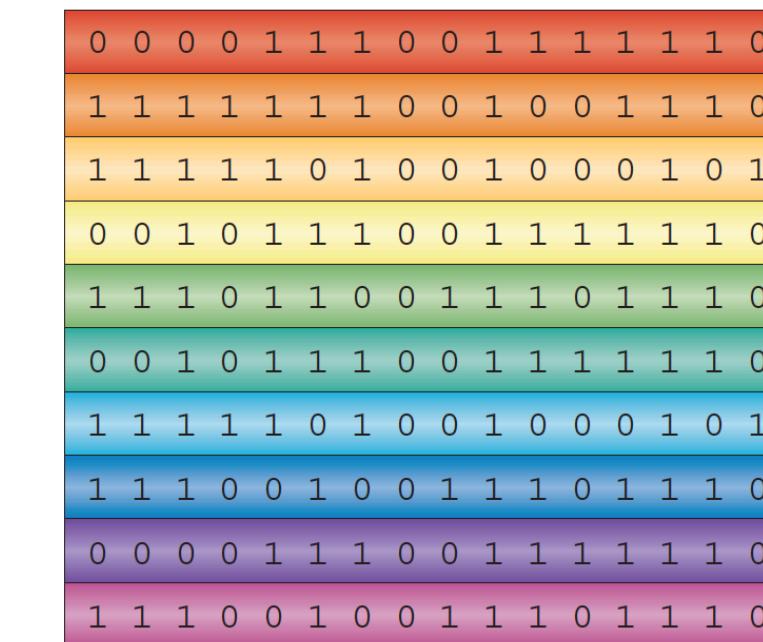
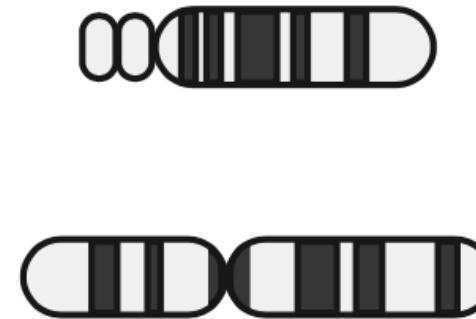


facultade de  
informática  
da coruña



Center for Research  
in Molecular Medicine and  
Chronic Diseases

# Comparación de la imputación genética en cohortes individuales *versus* imputación conjunta: Factores implicados y propuesta de actuación



Pedro Sánchez García

Directores Académicos: Raquel Cruz Guerrero y Silvia Diz de Almeida

Tutor Académico: Jorge González Domínguez



**Turno de preguntas**