



# SNP Markers for Quality Control in Cassava

Xiaofei Zhang, Luis Augusto Becerra, Adriana Bohorquez, Diana Victoria Marin, Tatiana Melissa Ovalle Rivera, Ismail Rabbi



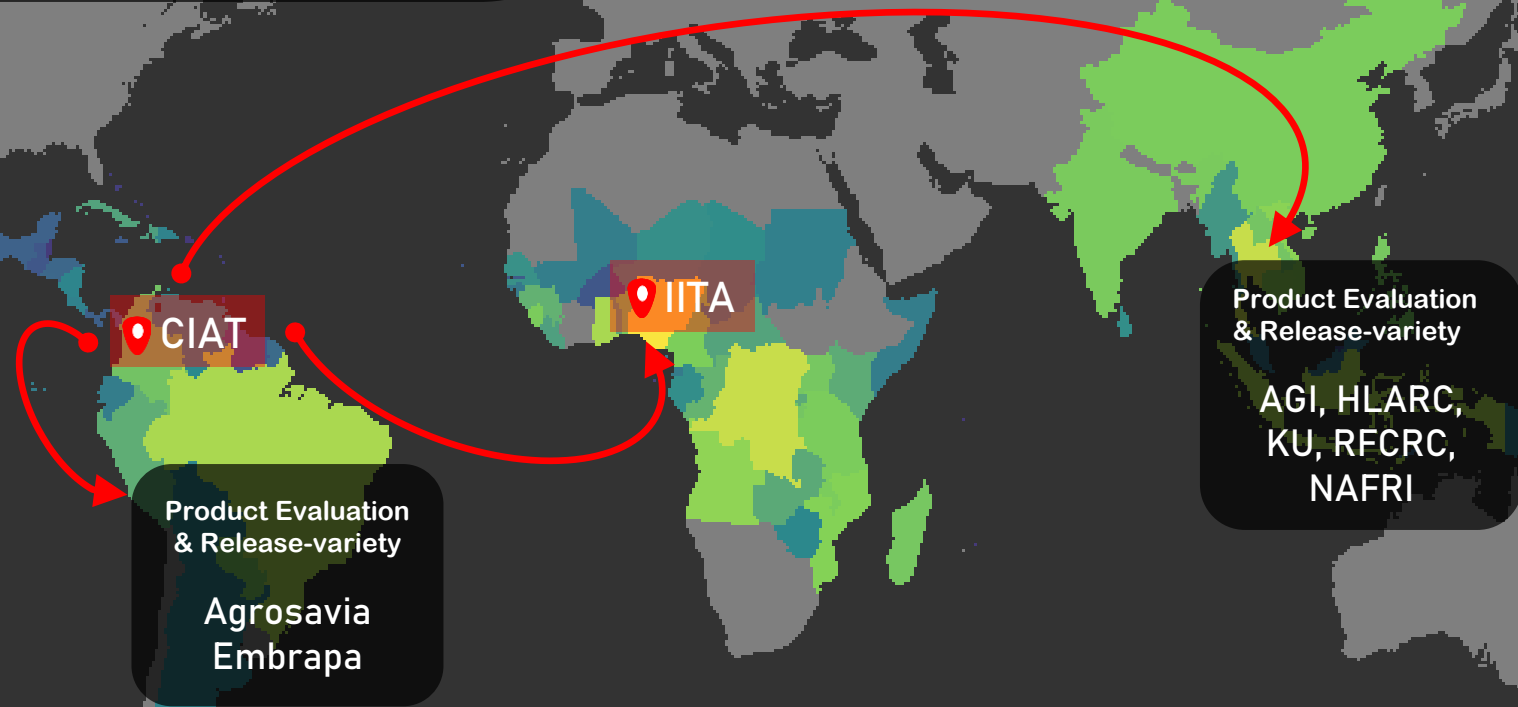
RESEARCH  
PROGRAM ON  
Roots, Tubers  
and Bananas



Alliance



Provide both trait donors and improved breeding populations with CMD, CBSD, and whitefly resistance to IITA and NARS.



Alliance

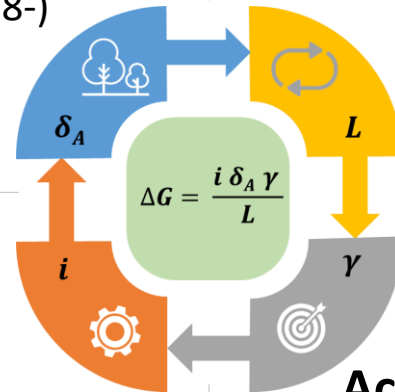
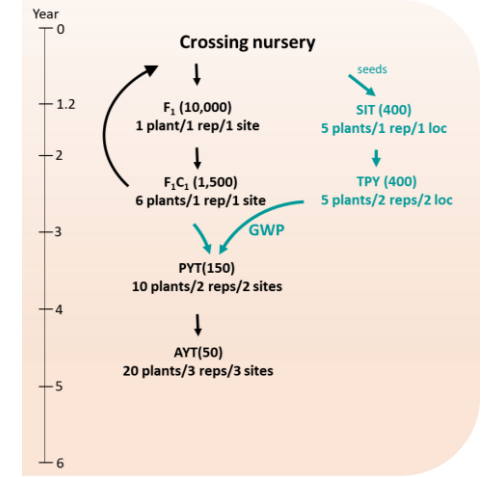
**Bioversity International**  
International Center for Tropical Agriculture  
Since 1977, Science to cultivate change

## Genetic Diversity

- New traits, e.g., CBSD res., CMD res., good cooking quality et al.
- Sequencing of progenitors (2020-)
  - Hybrid Breeding (2018-)

## Duration of Selection Cycle

- Flower Inducing (2016-)
- Genomewide Prediction (2019-)

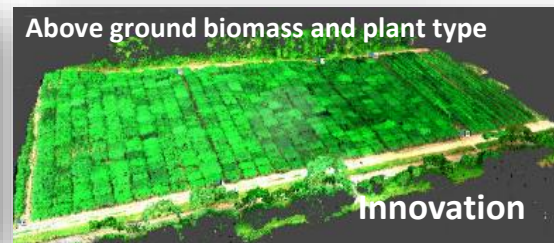
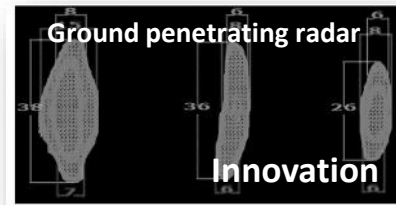


## Intensity

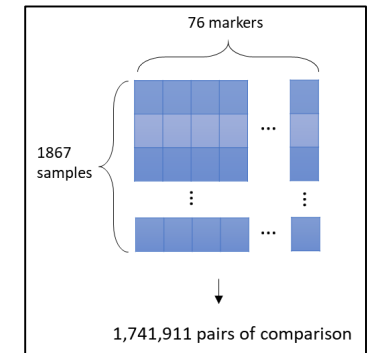
- High throughput phenotyping

## Accuracy

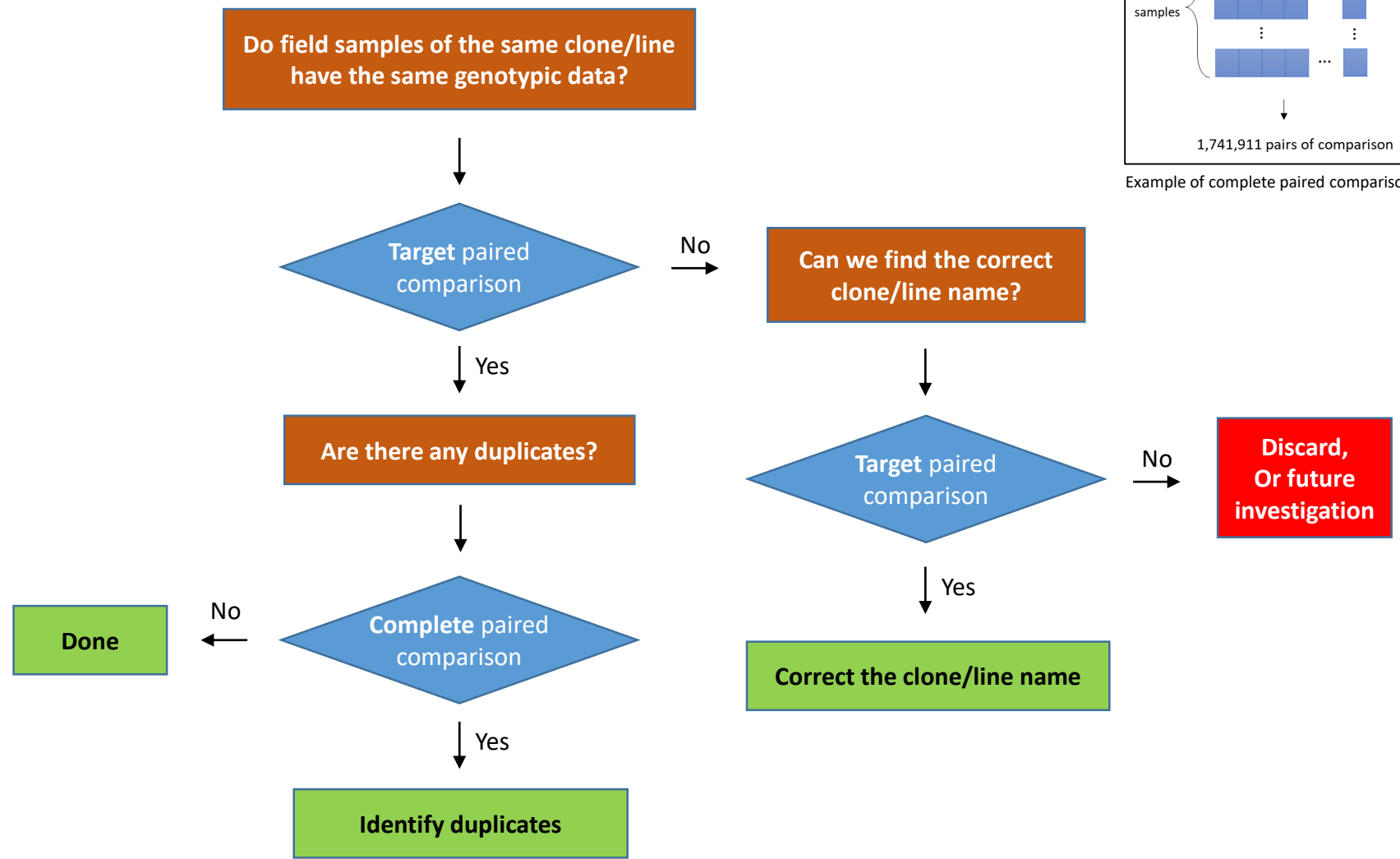
- CassavaBase, Fieldbook & Barcode (2018-)
- **Quality control** and MAS (2021-)
- TPE, ≥ 2 Environments (2020-)
- ≥ 5 Checks, BLUP and GBLUP (2020-)
- Selection Index (2012-)
- NIRS & Image Analysis (2012-)
- Stage&Gate System (2020-)
- Operational Excellence (2019-)



# Quality Control – Identity Checking



Example of complete paired comparison

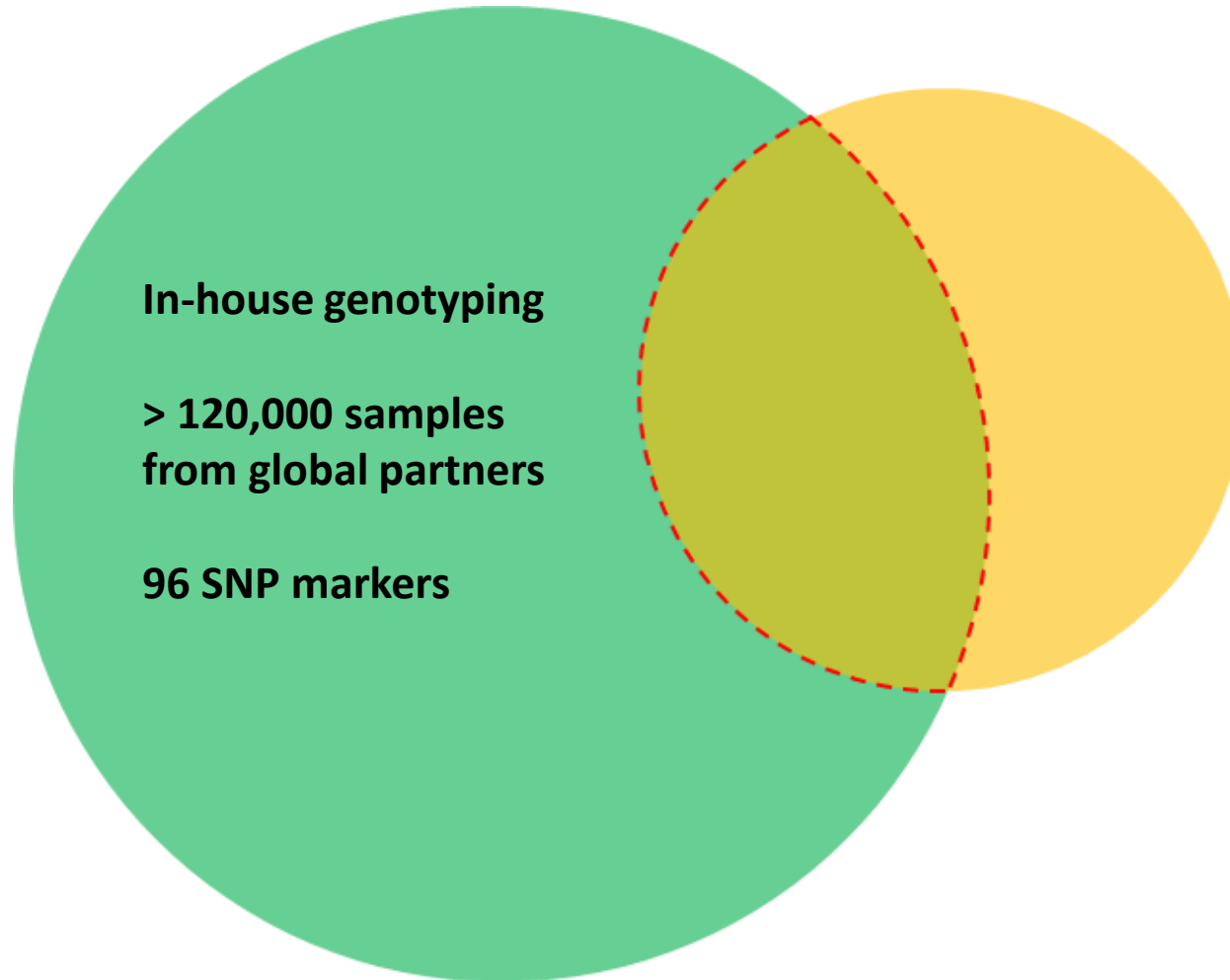




**Question:**

**What is the optimal set of markers?**

# Background



**In-house genotyping**

**> 120,000 samples  
from global partners**

**96 SNP markers**

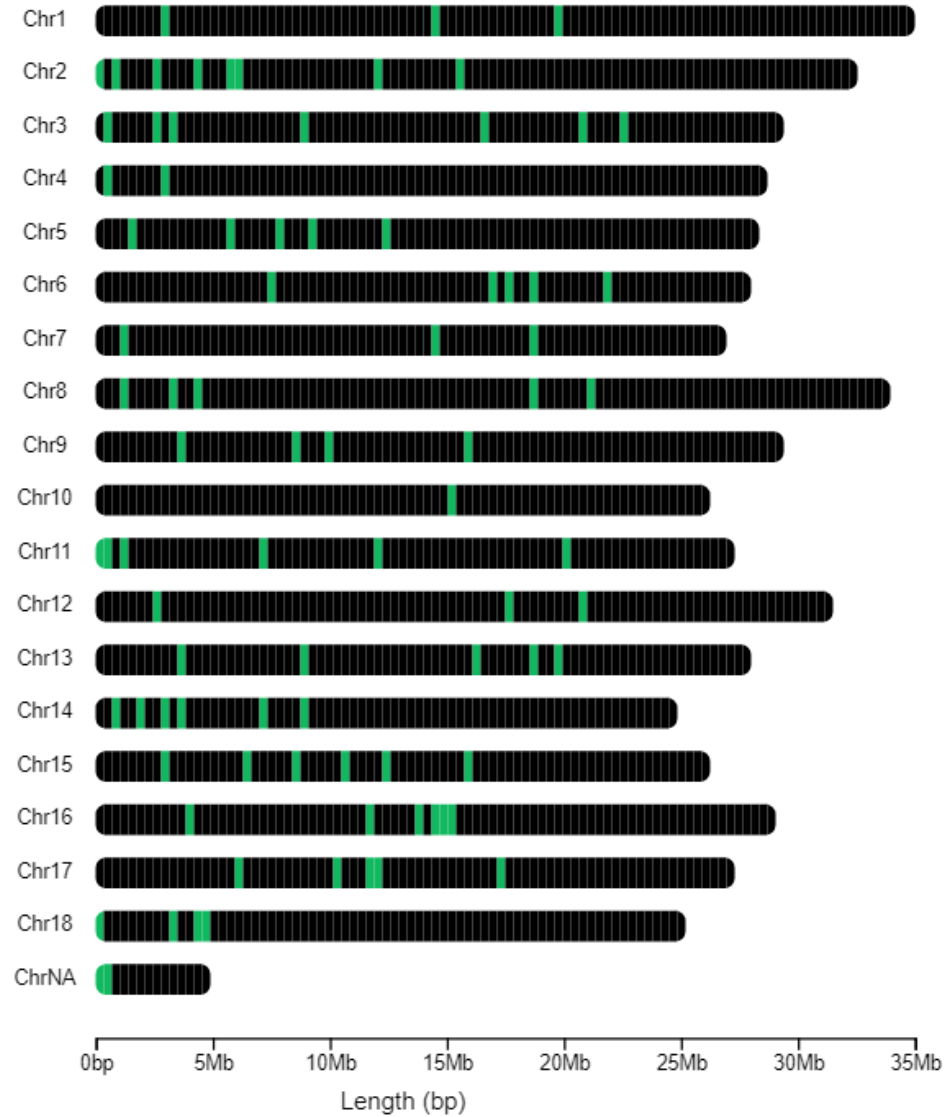
**Intertek genotyping**

**319 samples**

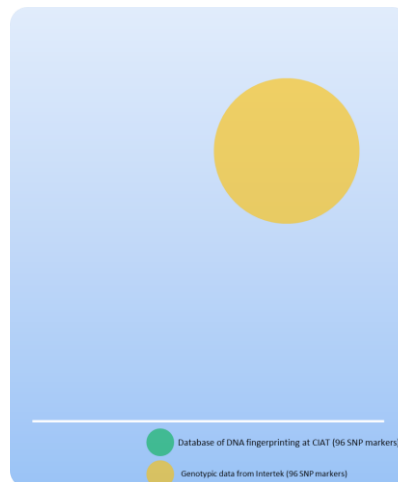
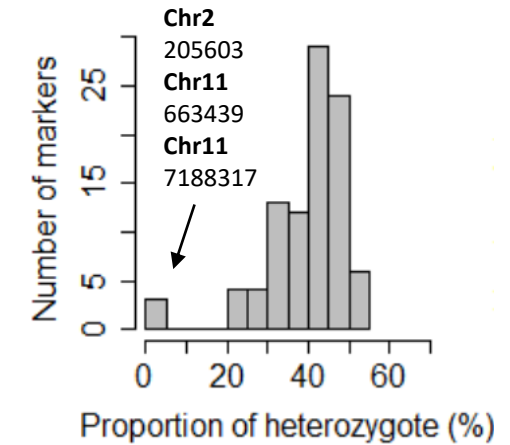
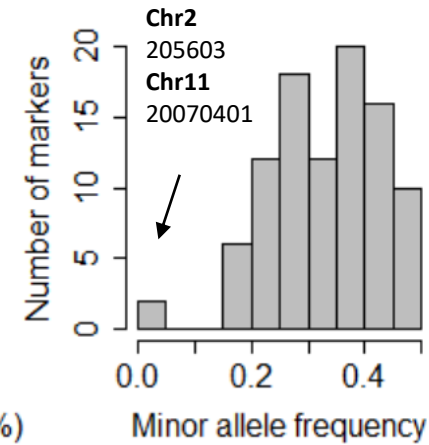
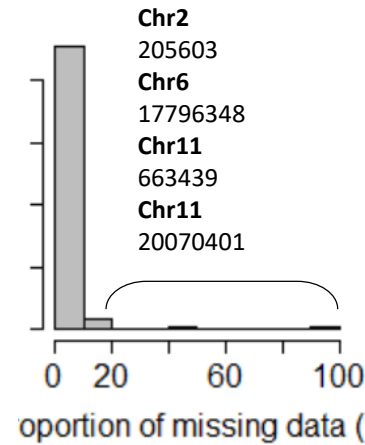
**96 SNP markers**

**DNA fingerprinting database at CIAT**

# 96 SNP Markers for DNA Fingerprinting



# Identify Problematic Markers in Intertek Platform



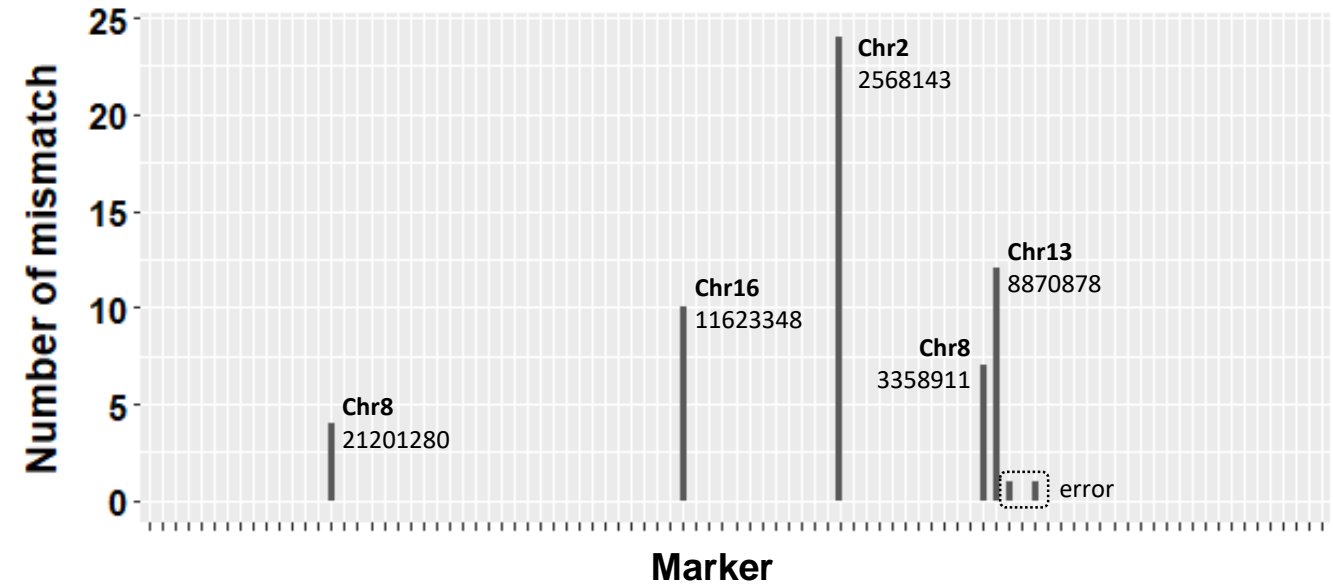
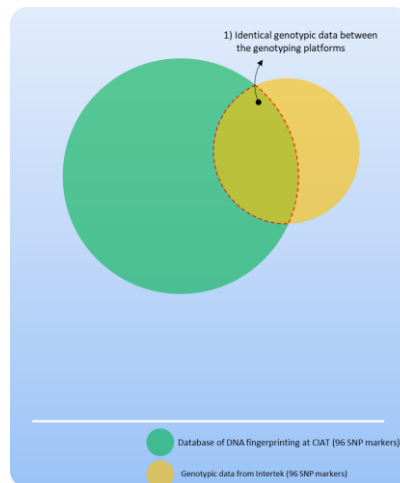
## Removed markers

Intertek	fluidigm	chr	pos
snpME00417	SNPY-016	6	17796348
snpME00426	SNPY-027	11	7188317
snpME00451	SNPY-086	11	20070401
snpME00479	SNPY-134	11	663439
snpME00501	SNPY-178	2	205603



# Compare Two Genotyping Platform

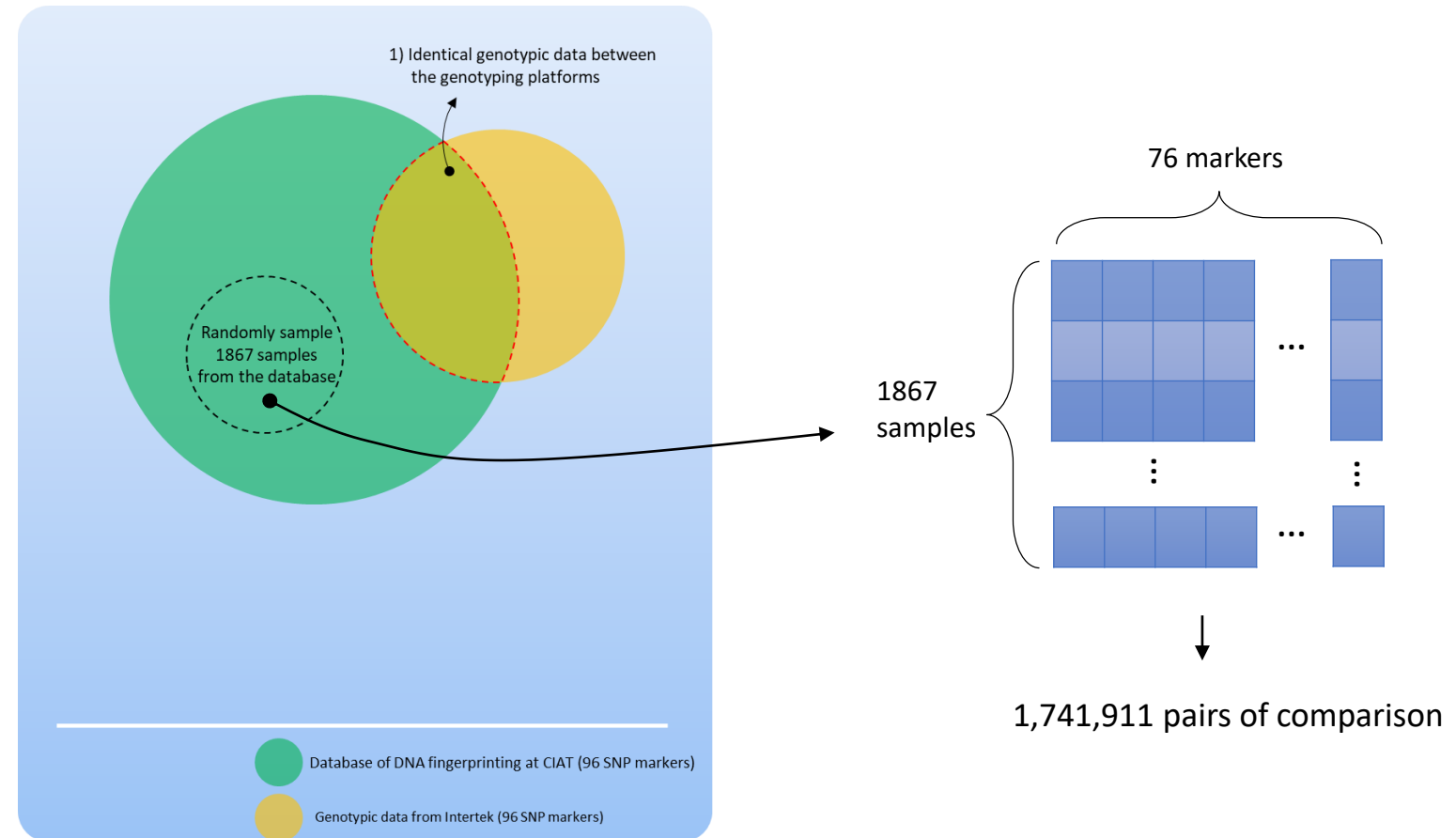
132 samples overlapped between the two genotyping platform



## Removed markers

Intertek	fluidigm	chr	pos
snpME00423	SNPY-024	8	21201280
snpME00452	SNPY-087	16	11623348
snpME00464	SNPY-109	2	2568143
snpME00475	SNPY-129	8	3358911
snpME00476	SNPY-130	13	8870878

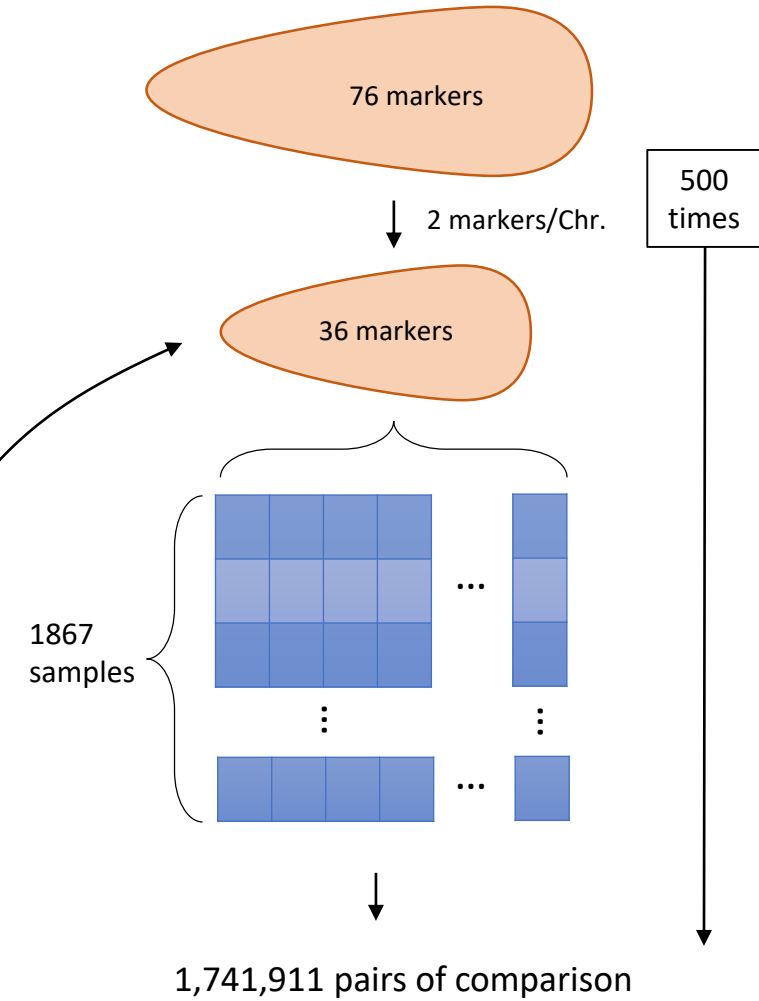
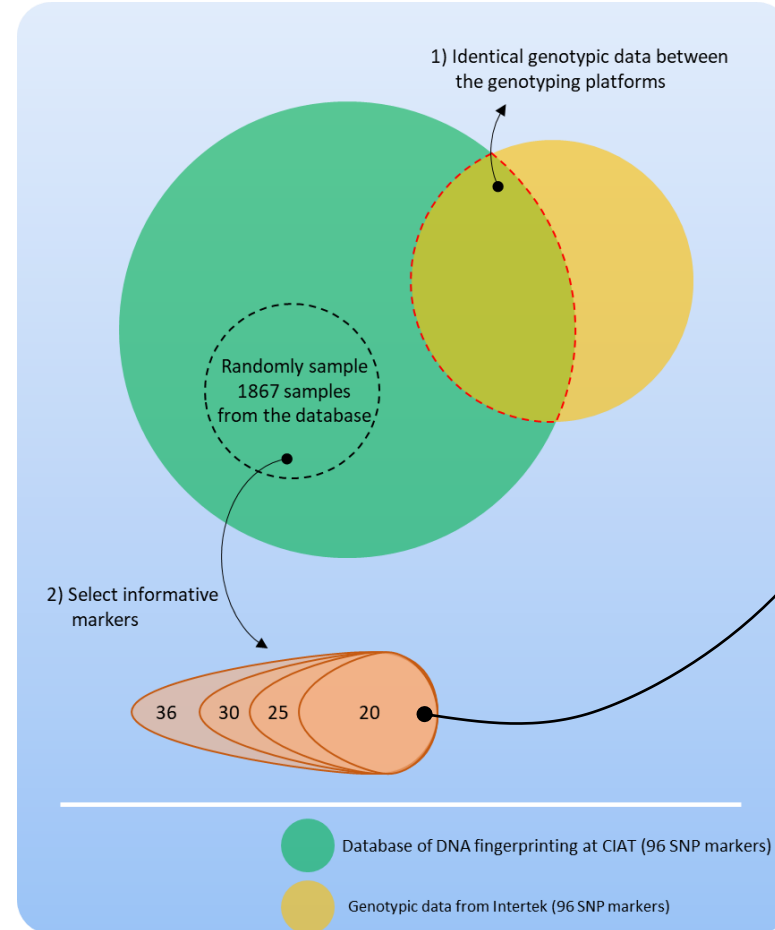
# Samples Used to Identify Informative Markers



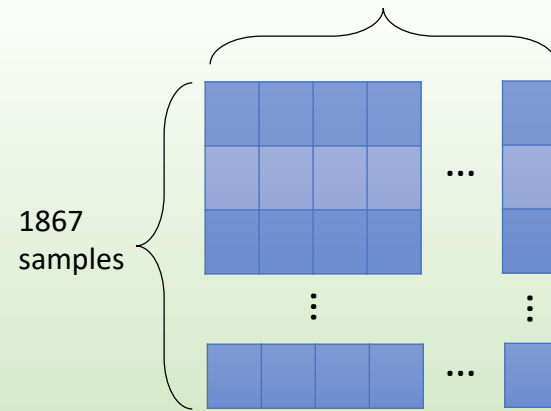
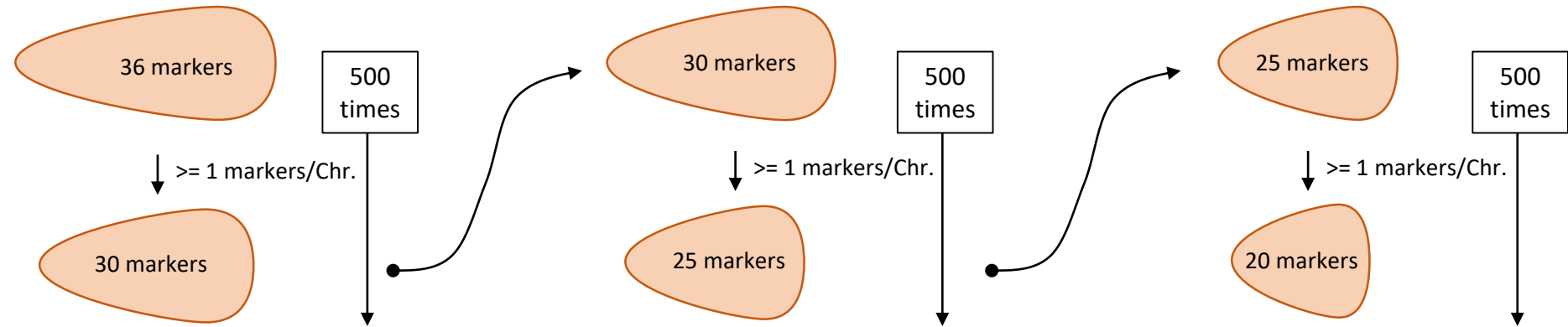
\* Remove 10 markers with more than 5% missing data

"SNPY-002" "SNPY-098" "SNPY-099" "SNPY-008" "SNPY-100" "SNPY-175" "SNPY-118" "SNPY-094" "SNPY-146" "SNPY-078"

# Determine the Informative Markers

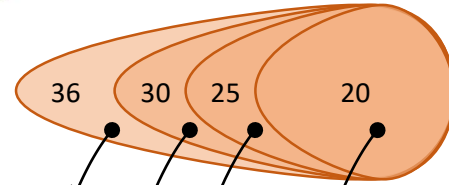


# Determine the Informative Markers



1,741,911 pairs of comparison

# The Optimum Number of Markers to Distinguish Samples



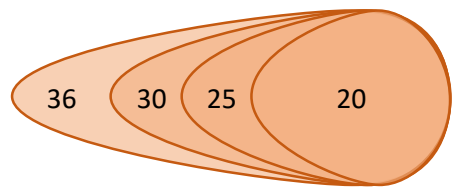
pair_ct_diff0	pair_ct_diff1	pair_ct_diff2	pair_ct_diff3	pair_ct_diff4	pair_ct_diff5	pair_ct_diff6	pair_ct_diff7
0	0	0	0	0	2	63	301

pair_ct_diff0	pair_ct_diff1	pair_ct_diff2	pair_ct_diff3	pair_ct_diff4	pair_ct_diff5	pair_ct_diff6	pair_ct_diff7
0	0	0	0	0	56	522	1398

pair_ct_diff0	pair_ct_diff1	pair_ct_diff2	pair_ct_diff3	pair_ct_diff4	pair_ct_diff5	pair_ct_diff6	pair_ct_diff7
0	0	0	0	227	903	2824	7758

pair_ct_diff0	pair_ct_diff1	pair_ct_diff2	pair_ct_diff3	pair_ct_diff4	pair_ct_diff5	pair_ct_diff6	pair_ct_diff7
0	0	5	192	1306	5596	18939	50272

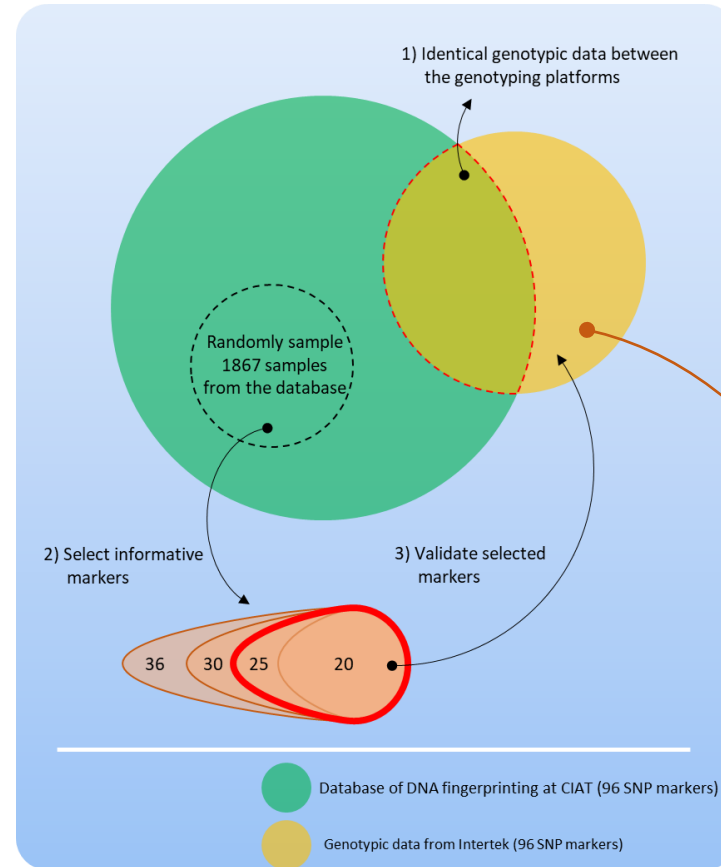
At least 3 markers different between any two samples



at least one marker from  
each chromosome

marker_CIAT	chr	36 markers	30 markers	25 markers	20 markers
SNPY-047	0	yes	yes	yes	yes
SNPY-048	1	yes	yes	yes	yes
SNPY-009	2	yes	yes	yes	yes
SNPY-080	3	yes	yes	yes	yes
SNPY-030	4	yes	yes	yes	yes
SNPY-164	5	yes	yes	yes	yes
SNPY-079	6	yes	yes	yes	yes
SNPY-066	7	yes	yes	yes	yes
SNPY-095	8	yes	yes	yes	yes
SNPY-053	9	yes	yes	yes	yes
SNPY-131	10	yes	yes	yes	yes
SNPY-127	11	yes	yes	yes	yes
SNPY-152	12	yes	yes	yes	yes
SNPY-161	13	yes	yes	yes	yes
SNPY-113	14	yes	yes	yes	yes
SNPY-142	14	yes	yes	yes	yes
SNPY-159	15	yes	yes	yes	yes
SNPY-165	16	yes	yes	yes	yes
SNPY-012	17	yes	yes	yes	yes
SNPY-026	18	yes	yes	yes	yes
SNPY-017	3	yes	yes	yes	
SNPY-154	8	yes	yes	yes	
SNPY-046	13	yes	yes	yes	
SNPY-120	16	yes	yes	yes	
SNPY-015	18	yes	yes	yes	
SNPY-090	0	yes	yes		
SNPY-004	4	yes	yes		
SNPY-011	6	yes	yes		
SNPY-105	7	yes	yes		
SNPY-149	11	yes	yes		
SNPY-179	2	yes			
SNPY-051	5	yes			
SNPY-137	9	yes			
SNPY-168	12	yes			
SNPY-096	15	yes			
SNPY-132	17	yes			

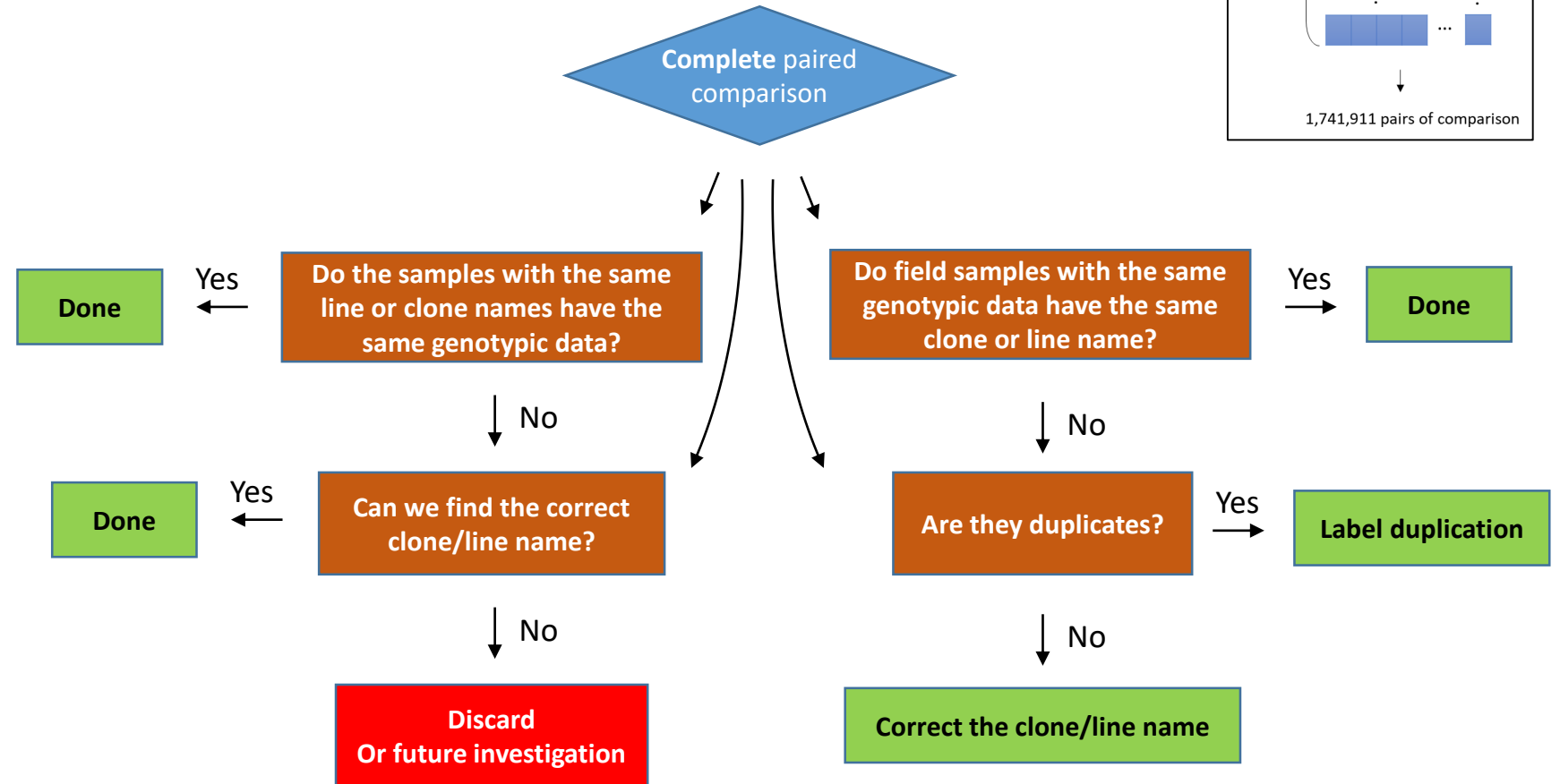
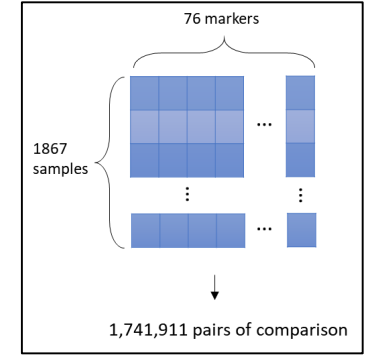
# Validate the 25 Markers in a Diversity Panel with 319 Genotypes



**Conclusion: The 25 markers are good to distinguish the 319 genotypes. At least 4 markers are different between any two genotypes**

pair_ct_diff0	pair_ct_diff1	pair_ct_diff2	pair_ct_diff3	pair_ct_diff4	pair_ct_diff5	pair_ct_diff6	pair_ct_diff7
0	0	0	0	0	8	38	165

# Quality Control – Identity Checking





**JOIN  
OUR TEAM**  
— BE PART OF OUR STORY —

[xiaofei.zhang@cigar.org](mailto:xiaofei.zhang@cigar.org)

Alliance



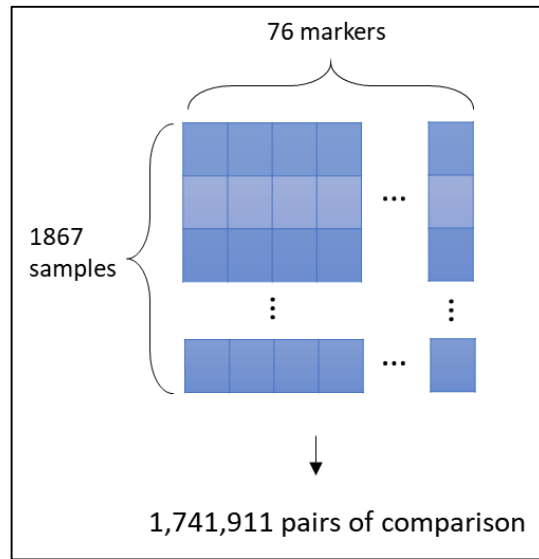
Recruiting: **Plant Geneticist**



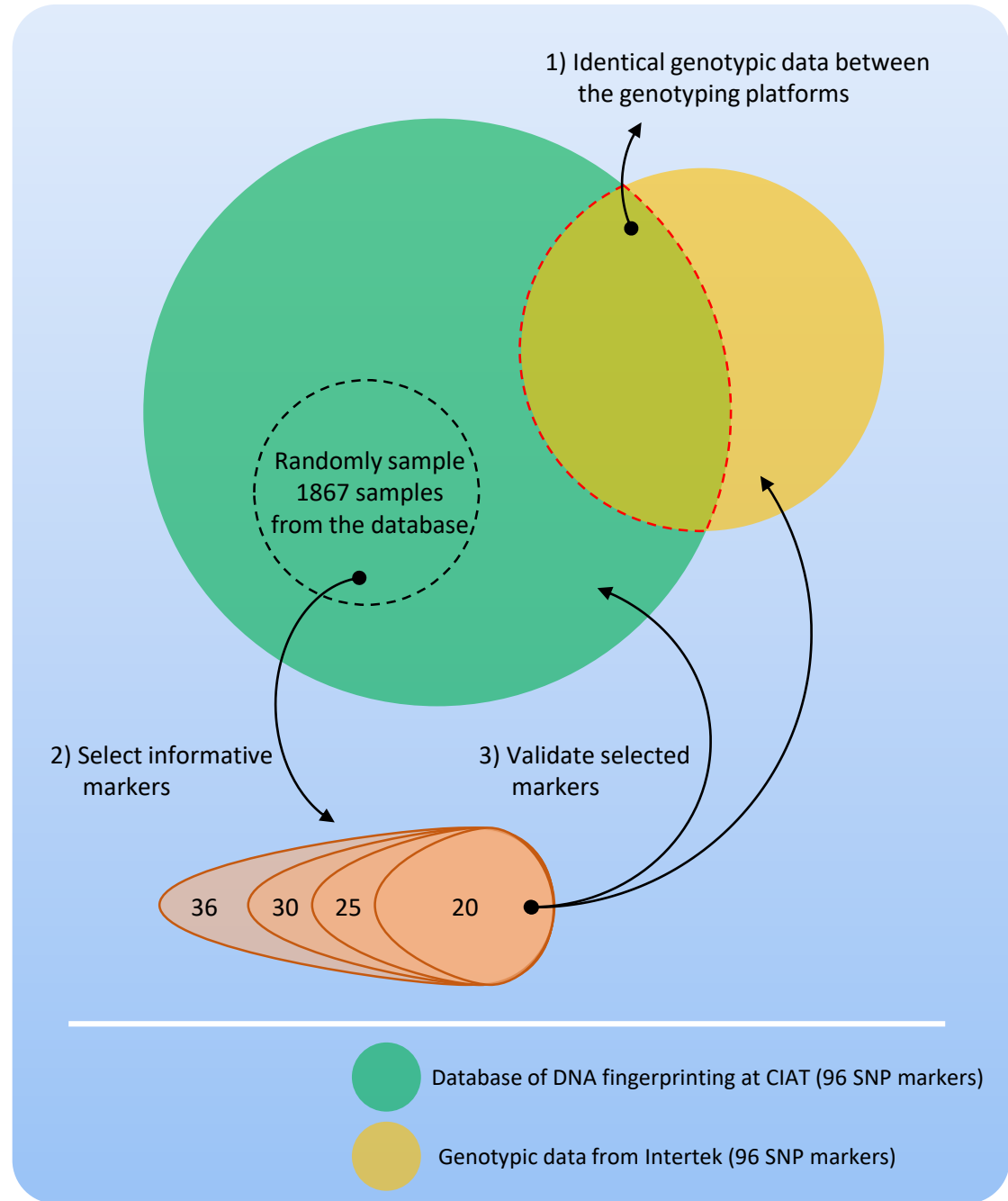




# Overview



Example of complete paired comparison



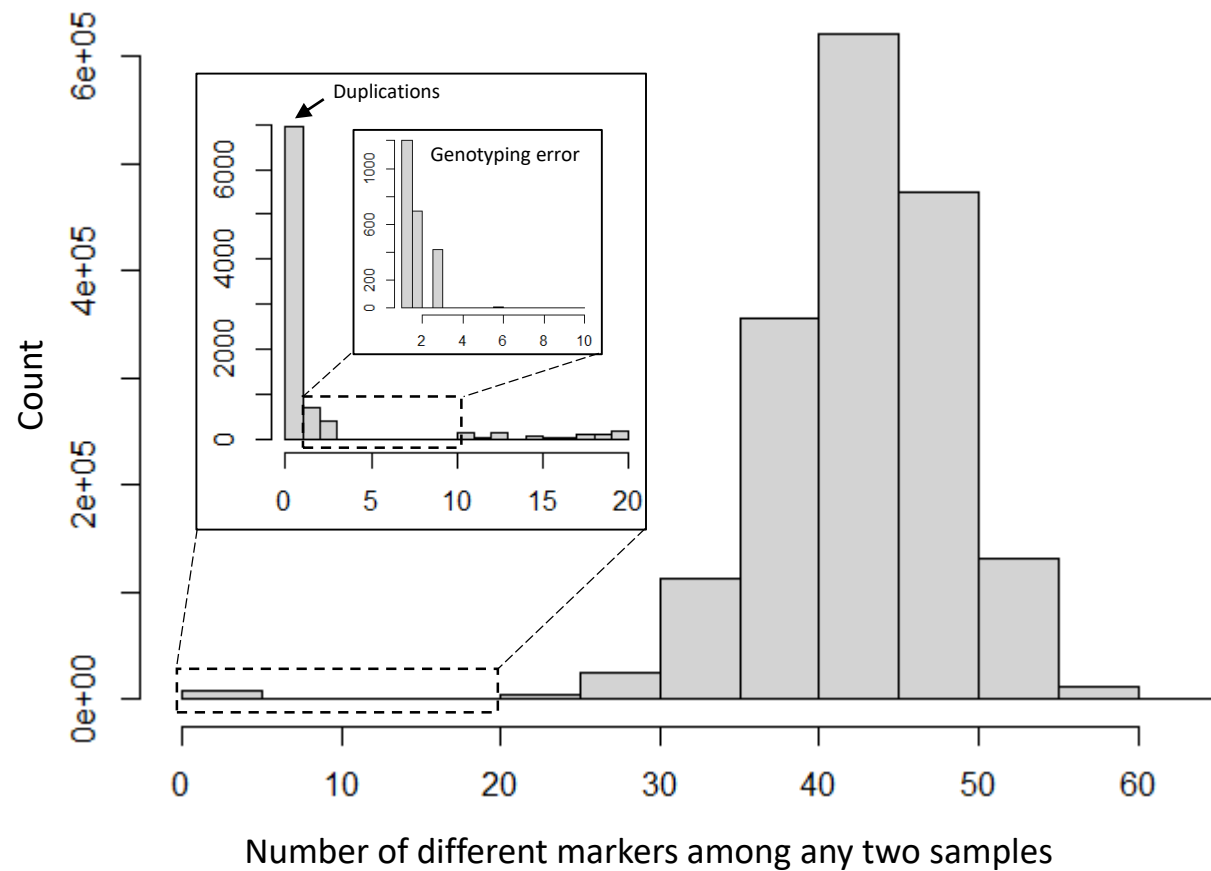
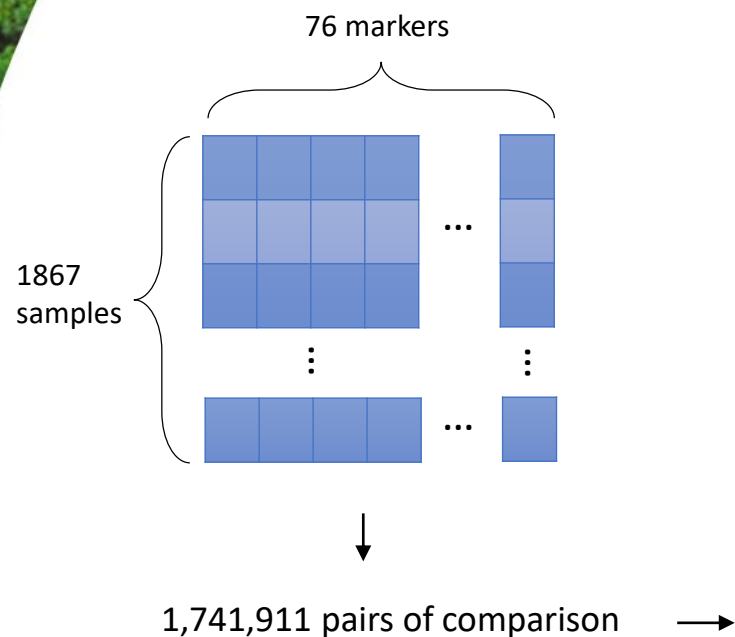


## More Than 19 Markers Different Between Any Two Genotypes

pair_ct_diff0	pair_ct_diff1	pair_ct_diff2	pair_ct_diff3	pair_ct_diff4	pair_ct_diff5	pair_ct_diff6	pair_ct_diff7
0	0	0	0	0	0	0	0
pair_ct_diff8	pair_ct_diff9	pair_ct_diff10	pair_ct_diff11	pair_ct_diff12	pair_ct_diff13	pair_ct_diff14	pair_ct_diff15
0	0	0	0	0	0	0	0
pair_ct_diff16	pair_ct_diff17	pair_ct_diff18	pair_ct_diff19	pair_ct_diff20	pair_ct_diff21	pair_ct_diff22	pair_ct_diff23
0	0	0	0	1	3	5	14
pair_ct_diff24	pair_ct_diff25	pair_ct_diff26	pair_ct_diff27	pair_ct_diff28	pair_ct_diff29	pair_ct_diff30	pair_ct_diff31
19	27	38	54	71	88	104	130
pair_ct_diff32	pair_ct_diff33	pair_ct_diff34	pair_ct_diff35	pair_ct_diff36	pair_ct_diff37	pair_ct_diff38	pair_ct_diff39
172	224	294	390	540	720	970	1277

\*91 SNP markers for 319 samples

# Difference Among 1867 Samples in 76 Markers\*



\* Remove 10 markers with more than 5% missing data

"SNPY-002" "SNPY-098" "SNPY-099" "SNPY-008" "SNPY-100" "SNPY-175" "SNPY-118" "SNPY-094" "SNPY-146" "SNPY-078"