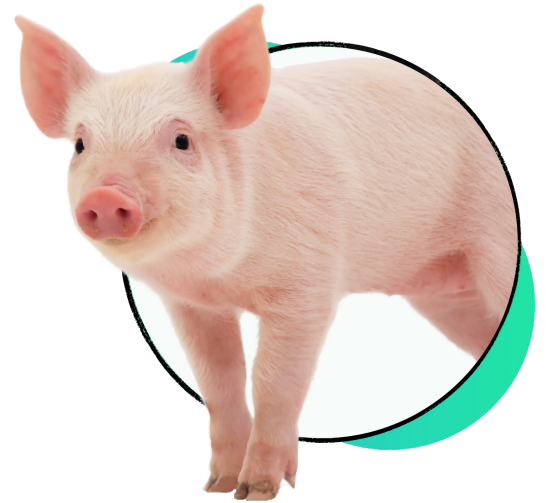


Определение породности свиней по данным генотипирования



Что такое геномная селекция?



Профиль SNP

Оценки племенной ценности

Подтверждение родства

Определение породности

Маркеры продуктивности

Генетические заболевания

Геномный инбридинг

Все это в ДНК!

Геномная селекция (*genomic selection*) – метод селекции, при котором оценка животных производится на основании определения большого количества однонуклеотидных замен (single nucleotide polymorphism, SNP).

Исходная информация

ДНК - микрочип



51 000 SNP

Аллели на гомологичных
хромосомах

SNP Name	Sample ID	Allele1 - Forward	Allele2 - Forward	Allele1 - AB	Allele2 - AB
1_10673082	piglet1	T	C	A	B
1_10723065	piglet1	T	T	A	A
1_11407894	piglet1	G	G	B	B
1_11426075	piglet1	T	T	A	A

51 000 строчек, которые содержат три варианта: AA (1/1), BB (2/2), AB (2/1).
Достаточное количество животных (> 100), принадлежащих к 3 породам.

[Simple guidelines for identifying top/bottom \(TOP/BOT\) strand and A/B allele.](#)

Частоты аллелей по Харди-Вайнбергу

Набор данных содержит J вариантов SNP у I животных, каждое из которых относится к K популяциям (породам).

Вероятность варианта AA в позиции j животного i

$$\Pr(1/1 \text{ for } i \text{ at SNP } j) = \left[\sum_k q_{ik} f_{kj} \right]^2$$

Вероятность варианта AB в позиции j животного i

$$\Pr(1/2 \text{ for } i \text{ at SNP } j) = 2 \left[\sum_k q_{ik} f_{kj} \right] \left[\sum_k q_{ik} (1 - f_{kj}) \right]$$

Вероятность варианта BB в позиции j животного i

$$\Pr(2/2 \text{ for } i \text{ at SNP } j) = \left[\sum_k q_{ik} (1 - f_{kj}) \right]^2.$$

f_{kj} - частота аллеля A в позиции j в породе k;

q_{ik} - доля генома k в генотипе животного i (породность).

Вычисление логарифмического правдоподобия

$$L(Q, F) = \sum_i \sum_j \left\{ g_{ij} \ln \left[\sum_k q_{ik} f_{kj} \right] + (2 - g_{ij}) \ln \left[\sum_k q_{ik} (1 - f_{kj}) \right] \right\}$$

g_{ij} Наблюдаемый маркер в позиции j животного i .

Условия:

$$0 \leq f_{kj} \leq 1, \quad q_{ik} \geq 0, \quad \sum_k q_{ik} = 1$$

желательно, чтобы
предполагаемый набор пород
соответствовал фактическому

Матрица $L(Q, F)$ - наблюдаемая вероятность аллельных комбинаций.

Необходимо подобрать параметры f_{kj} и q_{ik} такими, чтобы получить итоговую матрицу $L(Q, F)$.

David H. Alexander¹

John Novembre²

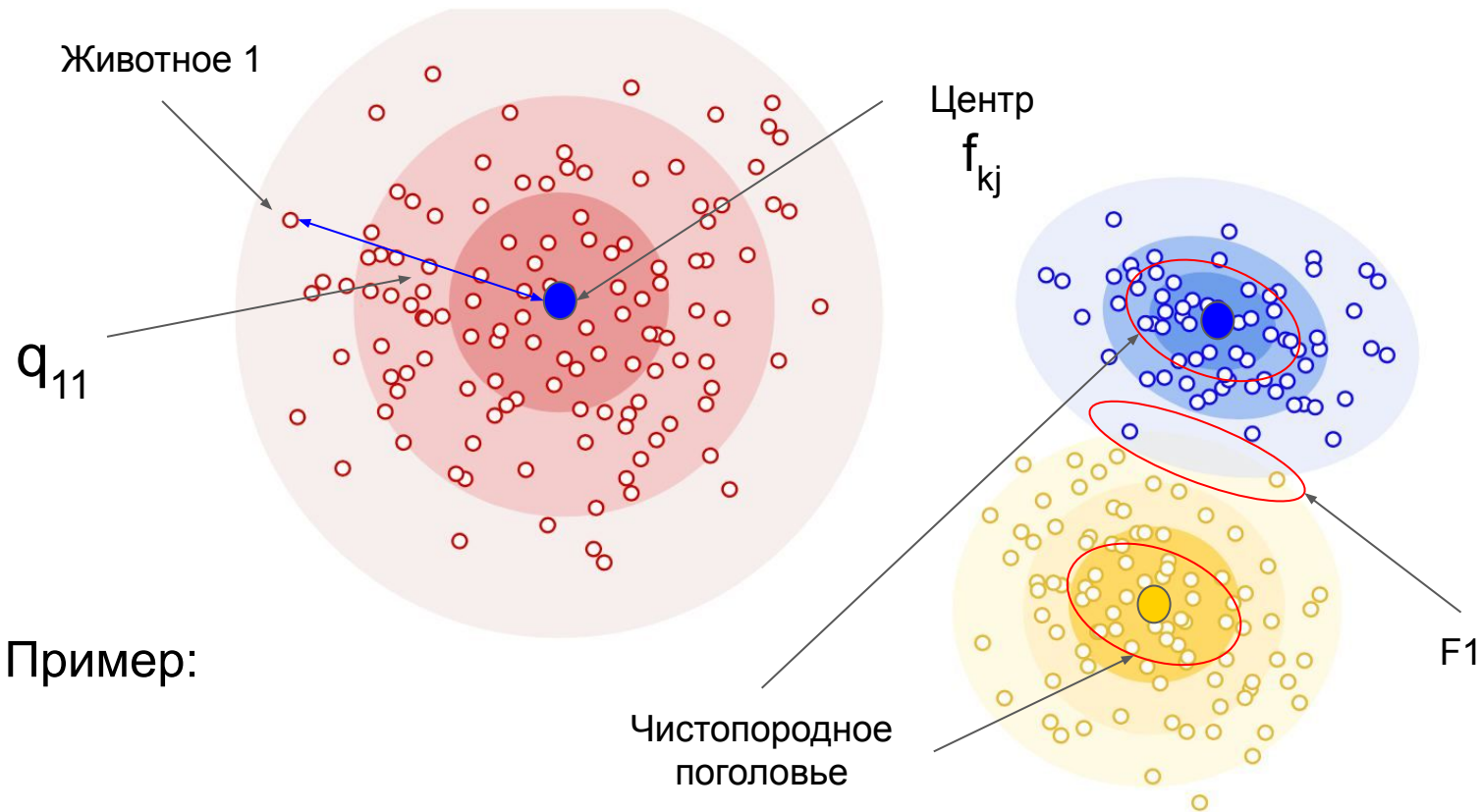
Kenneth Lange³

Fast Model-Based Estimation of Ancestry in Unrelated
Individuals

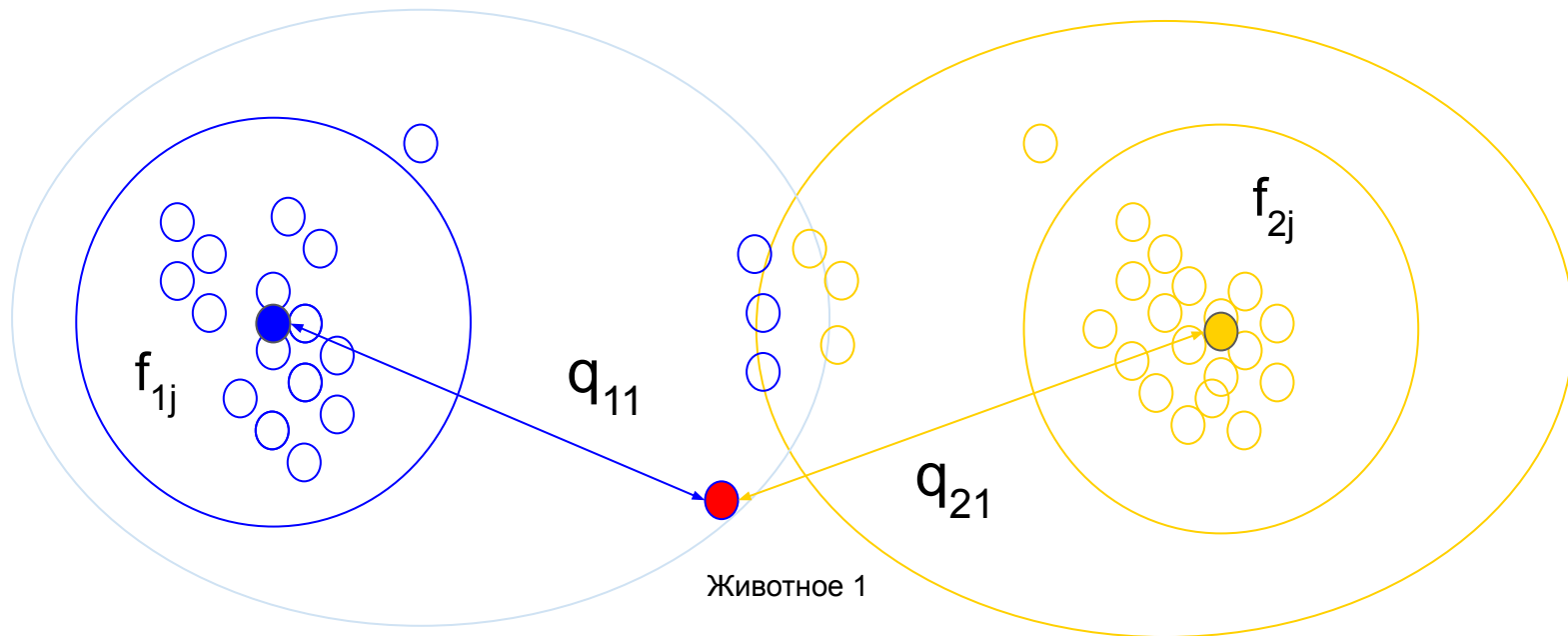
UCLA Department of Biomathematics.

Для подбора используются стандартные
математические методы выпуклой
оптимизации.

Определение породности свиней

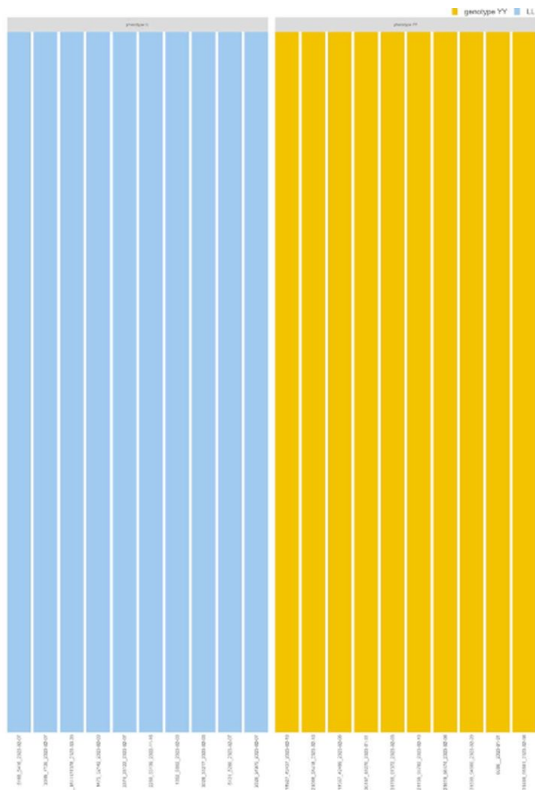


Определение породности свиней

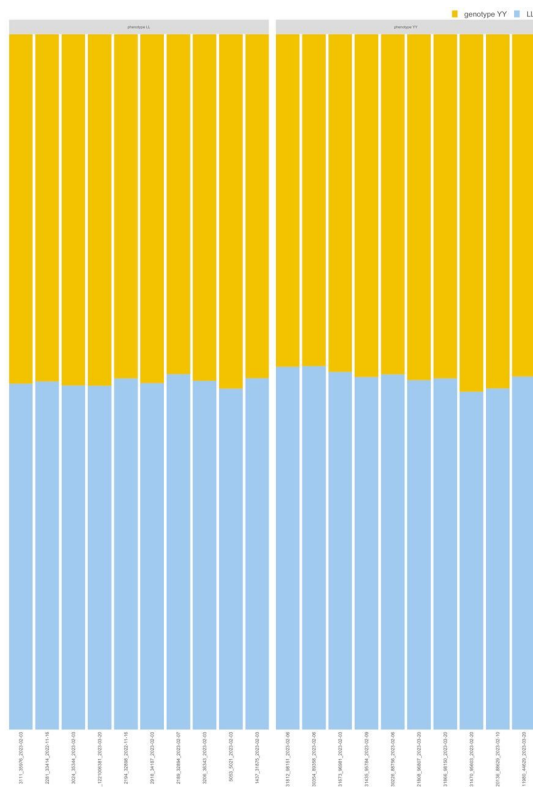


Номер животного <input type="text" value="1"/>	Породность Йоркшир <input type="text" value="46,3518%"/>	Породность Ландрас <input type="text" value="53,6482%"/>	Порода <input type="text" value="Йоркшир"/>
1	46,3518%	53,6482%	Йоркшир

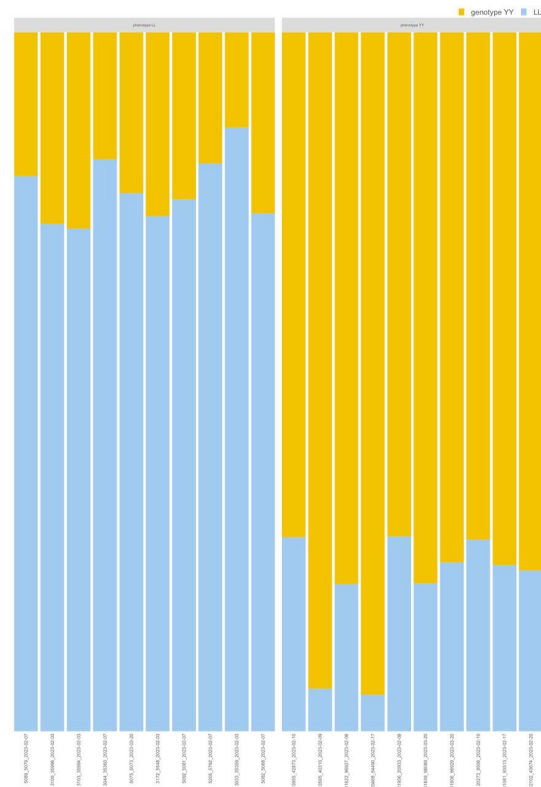
Результаты



Чистопородные животные

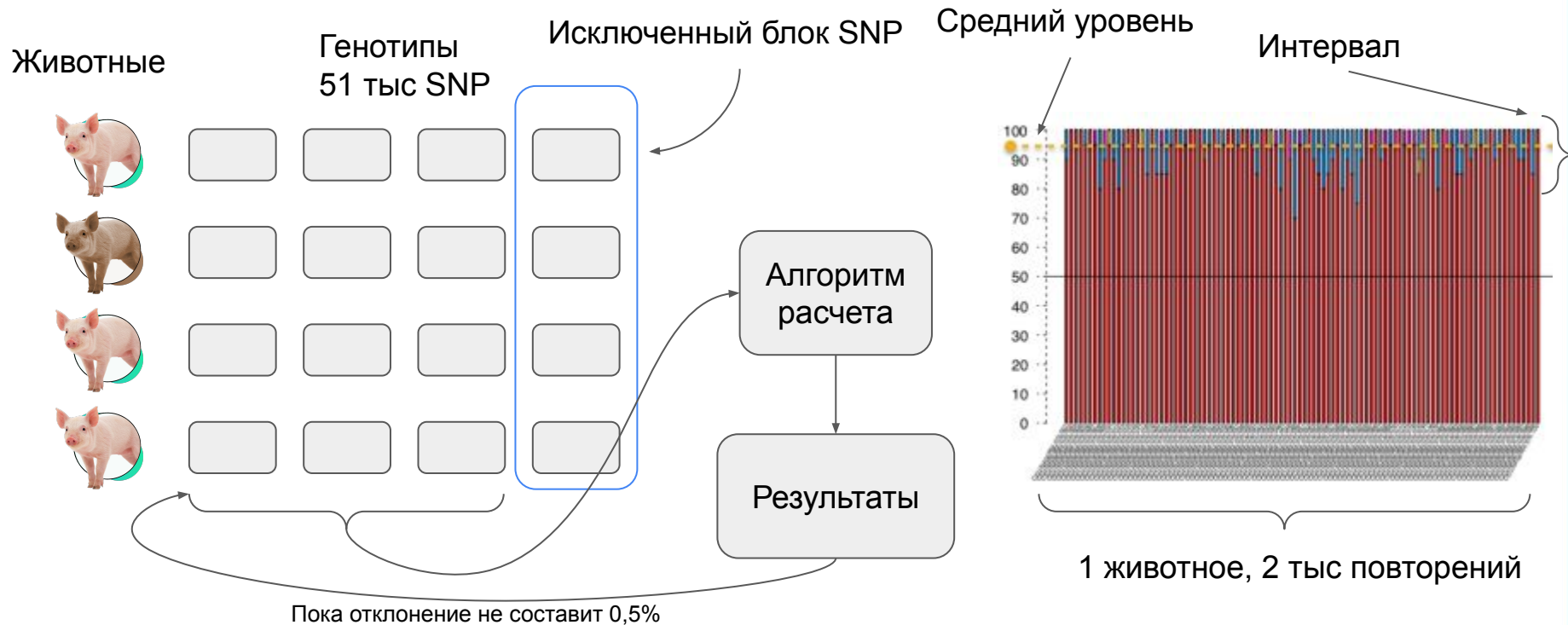


F1



Потомки F1

Расчет доверительного интервала

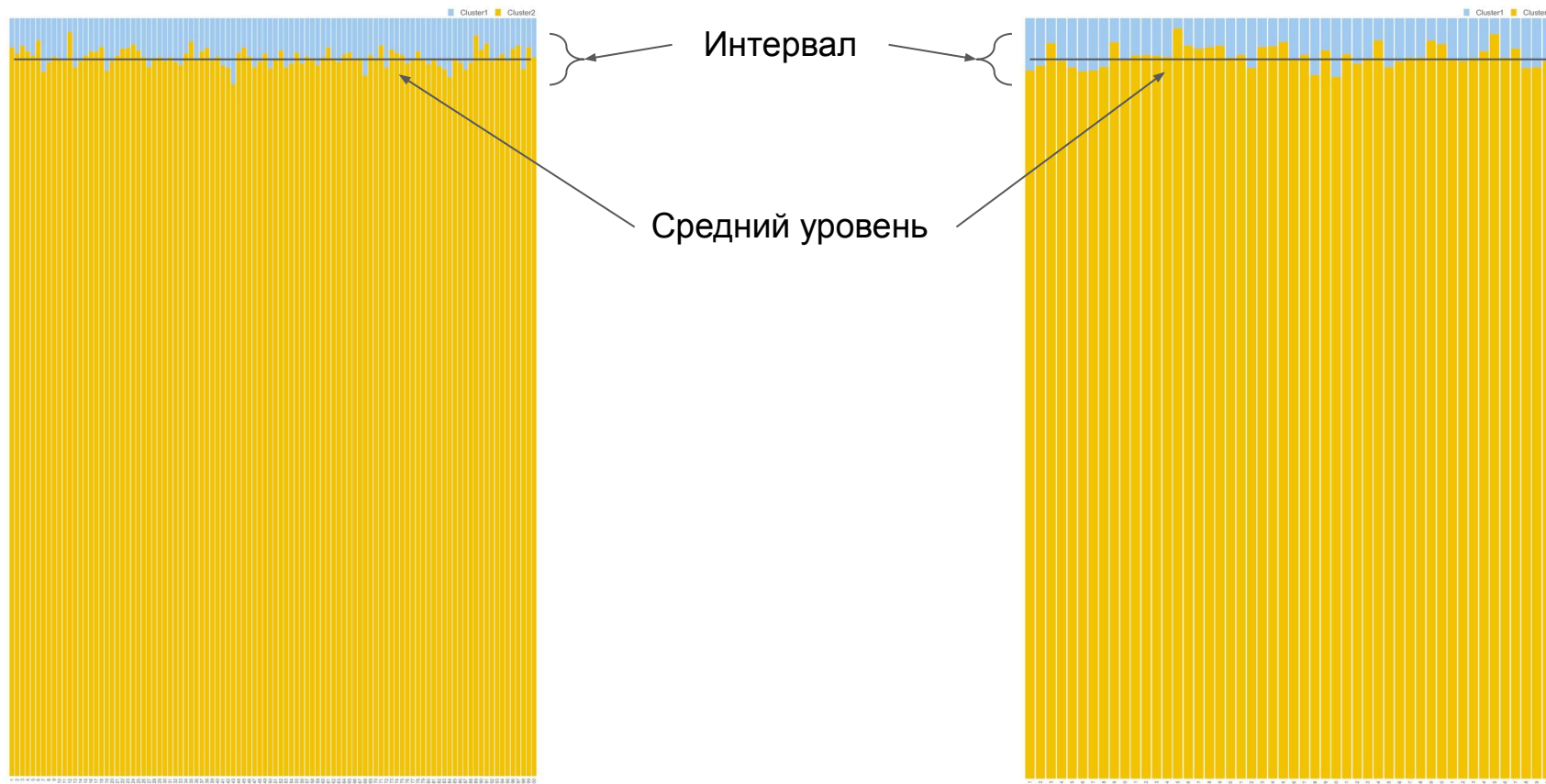


ANALYTICAL IMPROVEMENTS IN BIOGEOGRAPHIC ANCESTRY INFERENCE

M. Gabriel¹, C. Buchanan-Wright¹, A. Kumar¹, J. Lim¹, CW. Chang¹, J. Deng¹, R. Lagacé¹, S. Wootton¹

¹Thermo Fisher Scientific, South San Francisco, CA 94080

Расчет доверительного интервала



Метод блочной релаксации