Genetic Variants Regulating Immune Cell Levels in Health, Aging, and Disease

Siena 12 January 2016

Francesco Cucca

Institute for Genetic and Biomedical Research (IRGB), Italian National Research Council (CNR).

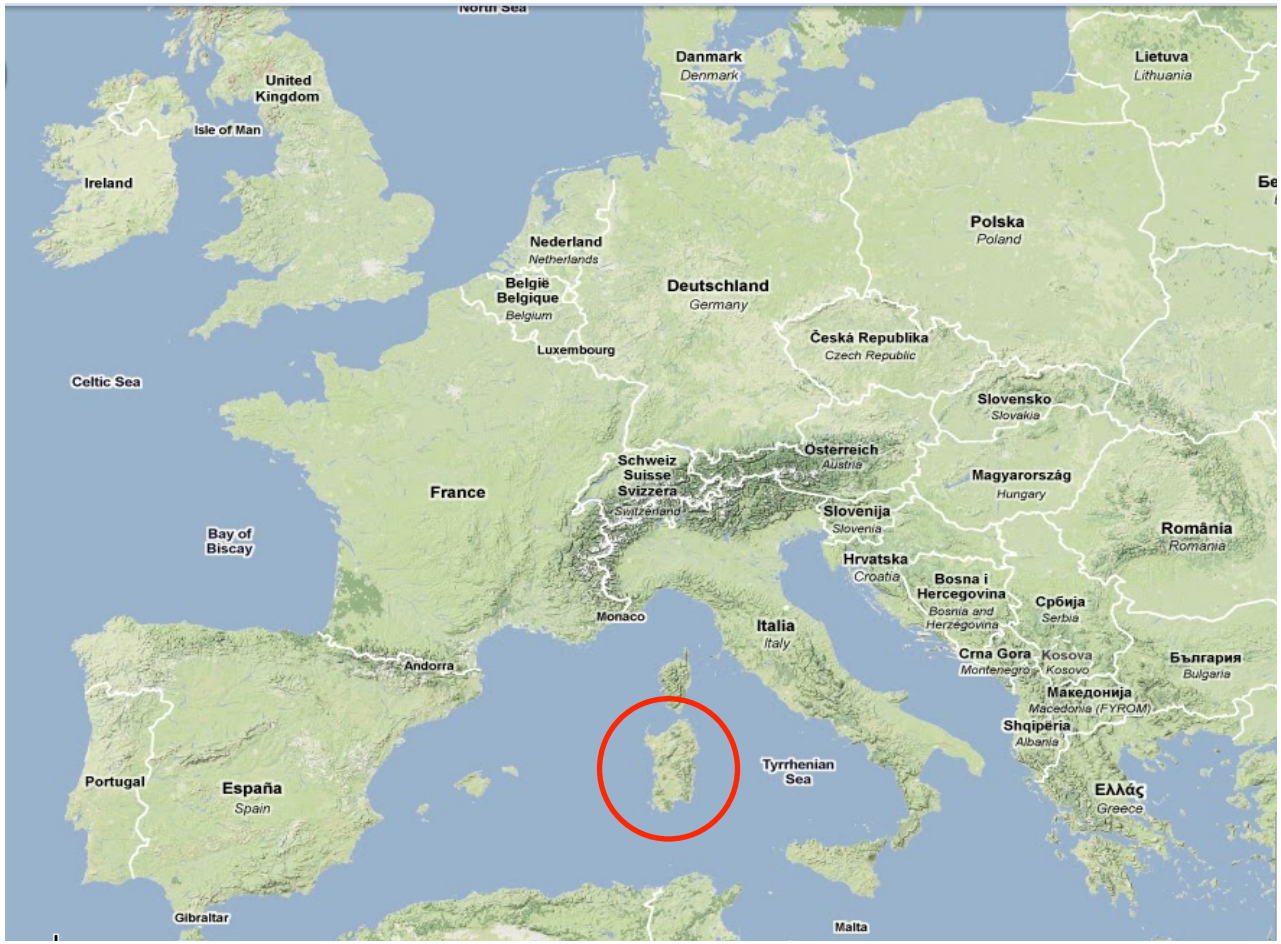
Chair of Medical Genetics, University of Sassari.

fcucca@uniss.it

Istituto di Ricerca Genetica e Biomedica

SardiNIA/progeNIA

Università di Sassari


North Sea
Danmark
Denmark
Lietuva
Lithuania
United Kingdom
Isle of Man
Ireland
Polska
Poland
Nederland
Netherlands
België
Belgique
Belgium
Deutschland
Germany
Luxembourg
Česká Republika
Czech Republic
Celtic Sea
Slovensko
Slovakia
Österreich
Austria
Schweiz
Suisse
Svizzera
Switzerland
France
Magyarország
Hungary
Slovenija
Slovenia
Bay of Biscay
România
Romania
Hrvatska
Croatia
Bosna i Hercegovina
Bosnia and Herzegovina
Србија
Serbia
Monaco
Italia
Italy
Crna Gora
Montenegro
Kosova
Kosovo
България
Bulgaria
Andorra
Македонија
Macedonia (FYROM)
Shqipëria
Albania
Tyrrhenian Sea
Portugal
España
Spain
Ελλάς
Greece
Gibraltar
Malta

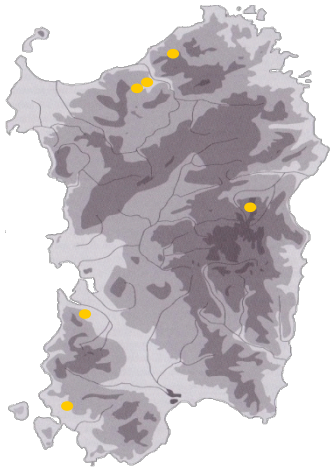
Archaeological evidence of human inhabitation of Sardinia

Radio-carbon dating archaeological sites

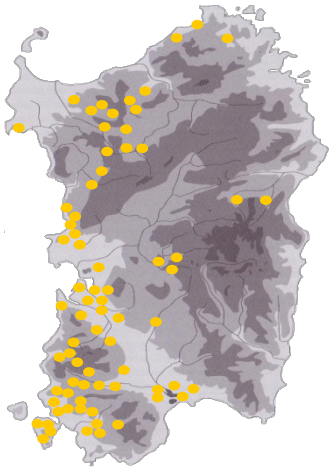
Before 7.7Kya few evidence of settlements

After 7.7Kya farming and breeding allowed expansion

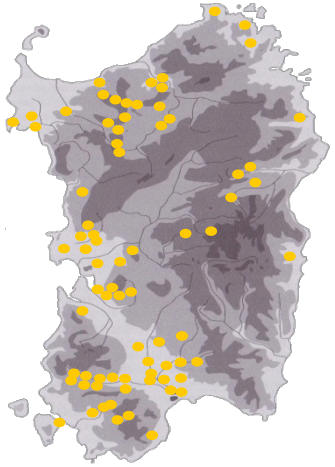
Large number of archaeological sites detected



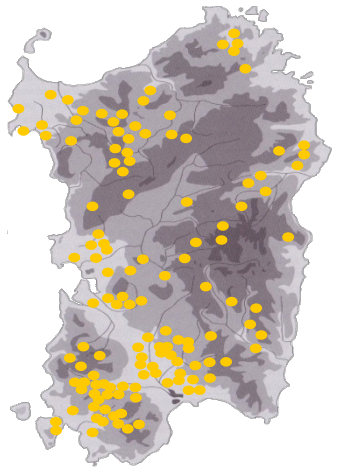
13-7.7KYA



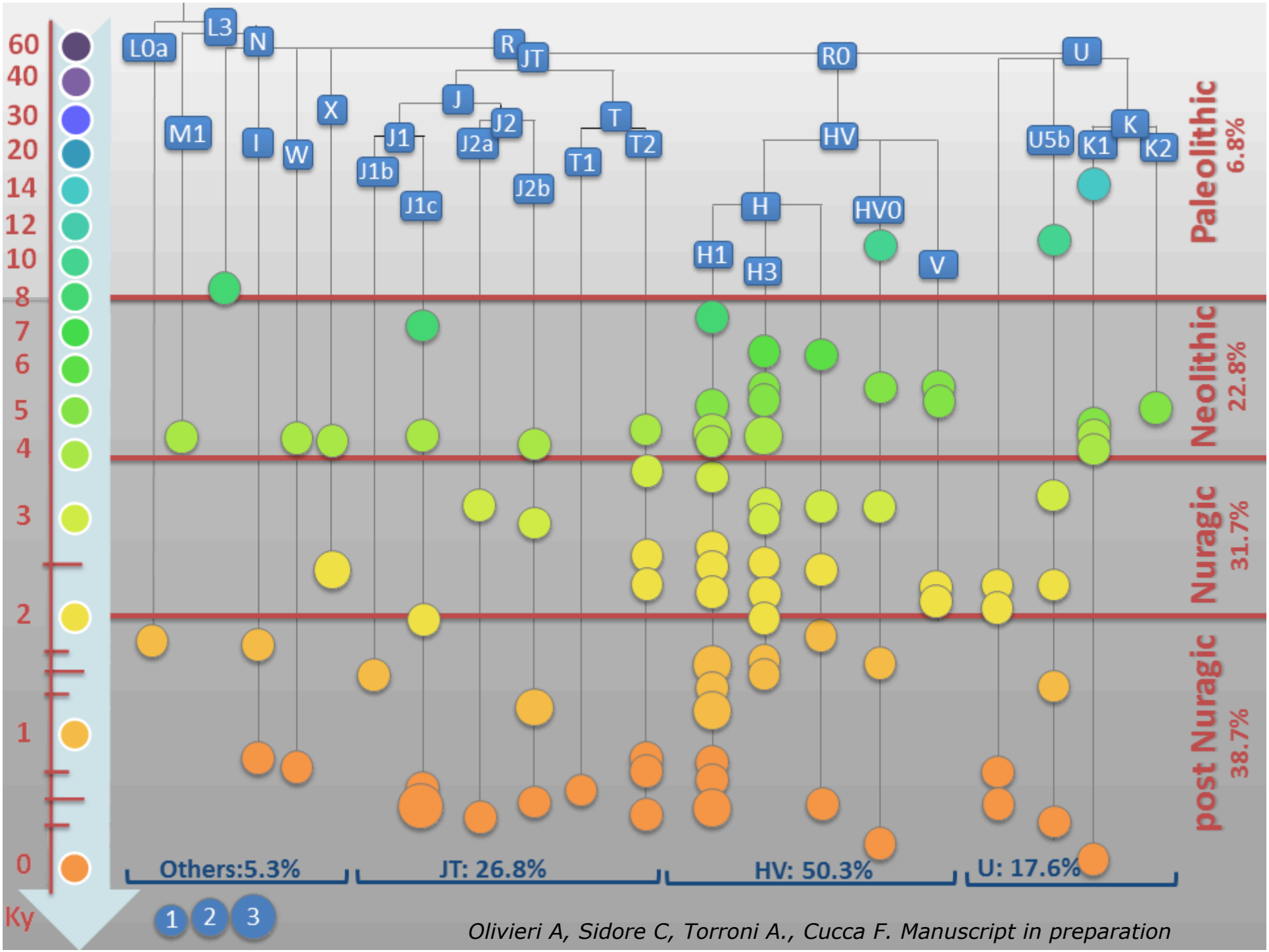
7.7-6.0KYA



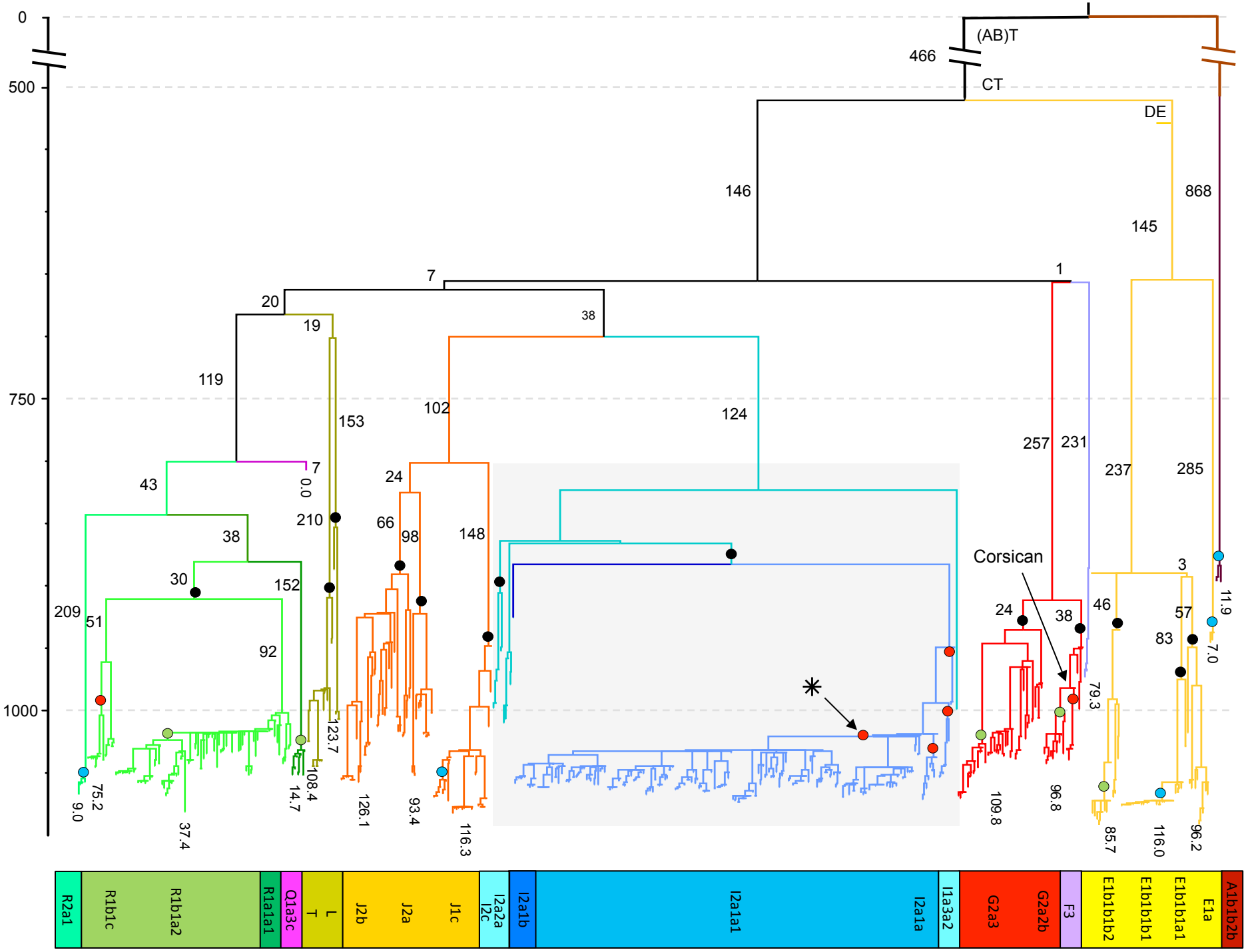
6.0-5.4KYA



5.4-4.8KYA



Olivieri A, Sidore C, Torroni A., Cucca F. Manuscript in preparation


0
500
750
1000
(AB)T
466
CT
DE
146
868
145
7
1
20
19
38
119
153
102
124
257
231
237
285
43
7
0.0
24
66
98
148
210
38
30
152
209
51
92
Corsican
3
46
24
38
57
83
11.9
7.0
79.3
123.7
108.4
14.7
75.2
9.0
37.4
126.1
93.4
116.3
109.8
96.8
85.7
116.0
96.2
R2a1
R1b1c
R1b1a2
R1a1a1
Q1a3c
L
T
J2b
J2a
J1c
I2a2a
I2c
I2a1b
I2a1a1
I2a1a
I1a3a2
G2a3
G2a2b
F3
E1b1b1b2
E1b1b1b1
E1b1b1a1
E1a
A1b1b2b

Low-Pass DNA Sequencing of 1200 Sardinians Reconstructs European Y-Chromosome Phylogeny

Paolo Francalacci,[1]* Laura Morelli,[1]† Andrea Angius,[2,3] Riccardo Berutti,[3,4] Frederic Reinier,[3] Rossano Atzeni,[3] Rosella Pilu,[2] Fabio Busonero,[2,5] Andrea Maschio,[2,5] Ilenia Zara,[3] Daria Sanna,[1] Antonella Useli,[1] Maria Francesca Urru,[3] Marco Marcelli,[3] Roberto Cusano,[3] Manuela Oppo,[3] Magdalena Zoledziewska,[2,4] Maristella Pitzalis,[2,4] Francesca Deidda,[2,4] Eleonora Porcu,[2,4,5] Fausto Poddie,[4] Hyun Min Kang,[5] Robert Lyons,[6] Brendan Tarrier,[6] Jennifer Bragg Gresham,[6] Bingshan Li,[7] Sergio Tofanelli,[8] Santos Alonso,[9] Mariano Dei,[2] Sandra Lai,[2] Antonella Mulas,[2] Michael B. Whalen,[2] Sergio Uzzau,[4,10] Chris Jones,[3] David Schlessinger,[11] Gonçalo R. Abecasis,[5] Serena Sanna,[2] Carlo Sidore,[2,4,5] Francesco Cucca[2,4]*

Genetic variation within the male-specific portion of the Y chromosome (MSY) can clarify the origins of contemporary populations, but previous studies were hampered by partial genetic information. Population sequencing of 1204 Sardinian males identified 11,763 MSY single-nucleotide polymorphisms, 6751 of which have not previously been observed. We constructed a MSY phylogenetic tree containing all main haplogroups found in Europe, along with many Sardinian-specific lineage clusters within each haplogroup. The tree was calibrated with archaeological data from the initial expansion of the Sardinian population ~7700 years ago. The ages of nodes highlight different genetic strata in Sardinia and reveal the presumptive timing of coalescence with other human populations. We calculate a putative age for coalescence of ~180,000 to 200,000 years ago, which is consistent with previous mitochondrial DNA–based estimates.

New sequencing technologies have provided genomic data sets that can reconstruct past events in human evolution more accurately (*1*). Sequencing data from the male-specific portion of the Y chromosome (MSY) (*2*), because of its lack of recombination and low mutation, reversion, and recurrence rates, can be particularly informative for these evolutionary analyses (*3*, *4*). Recently, high-coverage Y chromosome sequencing data from 36 males from different worldwide populations (*5*) assessed 6662 phylogenetically informative variants and estimated the timing of past events, including a putative coalescence time for modern humans of ~101,000 to 115,000 years ago.

MSY sequencing data reported to date still represent a relatively small number of individuals from a few populations. Furthermore, dating estimates are also affected by the calibration of the

[1]Dipartimento di Scienze della Natura e del Territorio, Università di Sassari, 07100 Sassari, Italy. [2]Istituto di Ricerca Genetica e Biomedica (IRGB), CNR, Monserrato, Italy. [3]Center for Advanced Studies, Research and Development in Sardinia (CRS4), Pula, Italy. [4]Dipartimento di Scienze Biomediche, Università di Sassari, 07100 Sassari, Italy. [5]Center for Statistical Genetics, Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, USA. [6]DNA Sequencing Core, University of Michigan, Ann Arbor, MI 48109, USA. [7]Center for Human Genetics Research, Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, TN 37235, USA. [8]Dipartimento di Biologia, Università di Pisa, 56126 Pisa, Italy. [9]Departamento de Genética, Antropología Física y Fisiología Animal, Universidad del País Vasco/Euskal Herriko Unibertsitatea, 48080 Bilbao, Spain. [10]Porto Conte Ricerche, Località Tramariglio, Alghero, 07041 Sassari, Italy. [11]Laboratory of Genetics, National Institute on Aging, Baltimore, MD 21224, USA.

*Corresponding author. E-mail: pfrancalacci@uniss.it (P.F.); fcucca@uniss.it (F.C.)

†Laura Morelli prematurely passed away on 20 February 2013. This work is dedicated to her memory.

Key genetic features of the Sardinians

- ✓ Large, ancient isolate
- ✓ Excellent depiction of European genetic variation
- ✓ Founder effects, genetic drift
- ✓ Strong Malaria pressure until its recent eradication
- ✓ Different allele frequencies
- ✓ Different LD patterns
- ✓ Affinity of contemporary Sardinian genomes with those of Early Neolithic farmers from across Europe

Key ingredients of the project :

- Selection of the most appropriate **phenotypes** in a **favorable study population**

- **Deep extraction** of the genetic information

Key ingredients of the project:

- Selection of the most appropriate **phenotypes** in a **favorable study population**

- **Deep extraction** of the genetic information

Projects ongoing

Study of Multiple Sclerosis & Type 1 Diabetes (CSCT)

~2000 T1D patients

~3000 MS patients

~4000 Controls

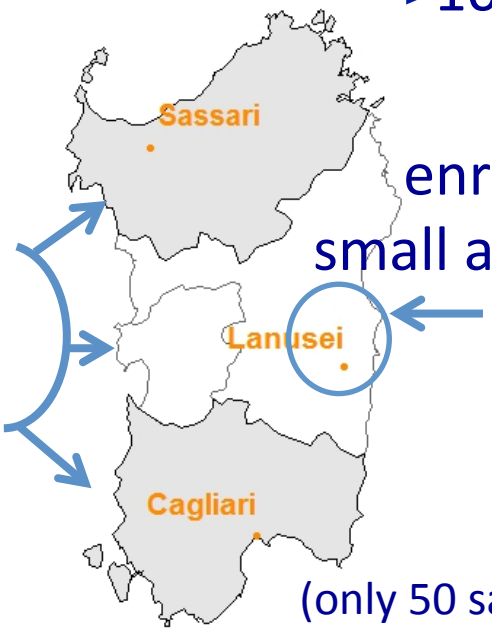
Affected individuals and matched controls from all over the island

SardiNIA general population cohort study on qunatitative

~7,000 individuals

1257 families

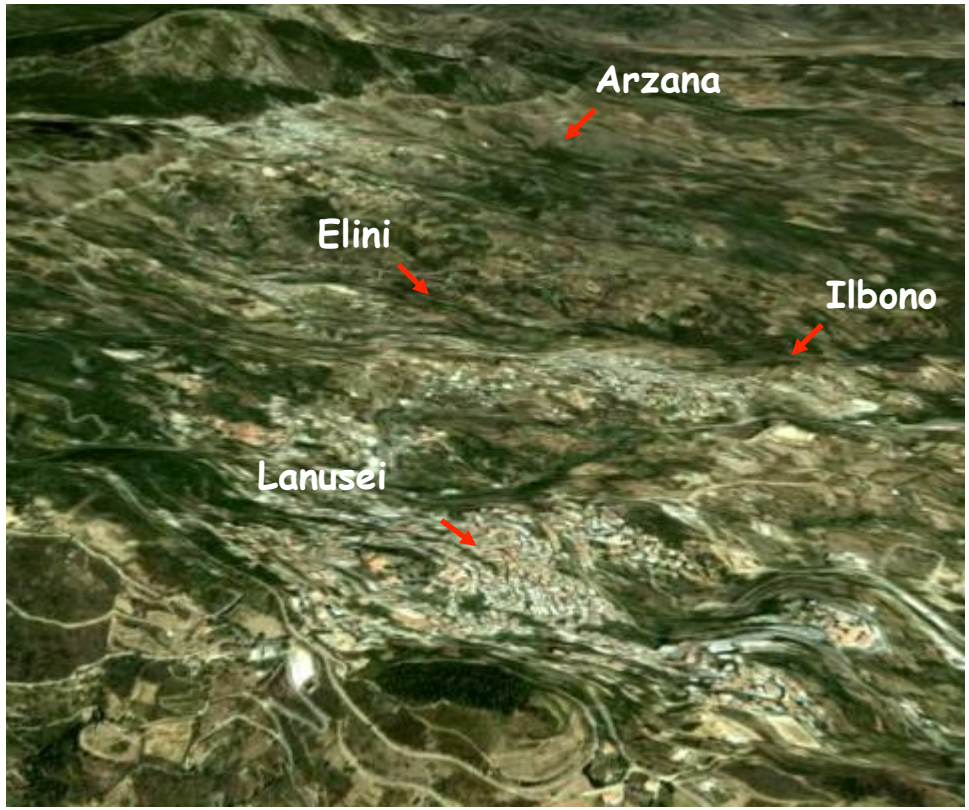
>1000 quantitative traits



enrollment from 4 towns in a small area of the Island

(only 50 samples enrolled in both studies)

The SardiNIA/ProgeNIA project



Rationale

- Focus on **geographically clustered towns** to permit extensive phenotyping with enough power to find associations with traits (mainly quantitative).

- The structure of the project allows **precise measures of heritability** and **simplifies imputation strategies**.

The SardiNIA/ProgeNIA project

- B**egan in 2001**, funded by the NIH/NIA to Giuseppe Pilia
- **~7,000 volunteers**, grouped in **~1,000 families** recruited from a cluster of **4 small towns** in the Ogliastra region
- **~1,500 quantitative traits** assessed thus far, with a strong emphasis on immune parameters and including anthropometric, haematological and cardiovascular measures.
- **Longitudinal study** – repeated (and new) measurements every ~3 years
- **~6,600 samples genotyped with 4 arrays** (OmniExpress, Immuno-, Metabo-, Exome- Illumina BeadChip)

Key ingredients of the project:

- Selection of the most appropriate **phenotypes** in the study **population**

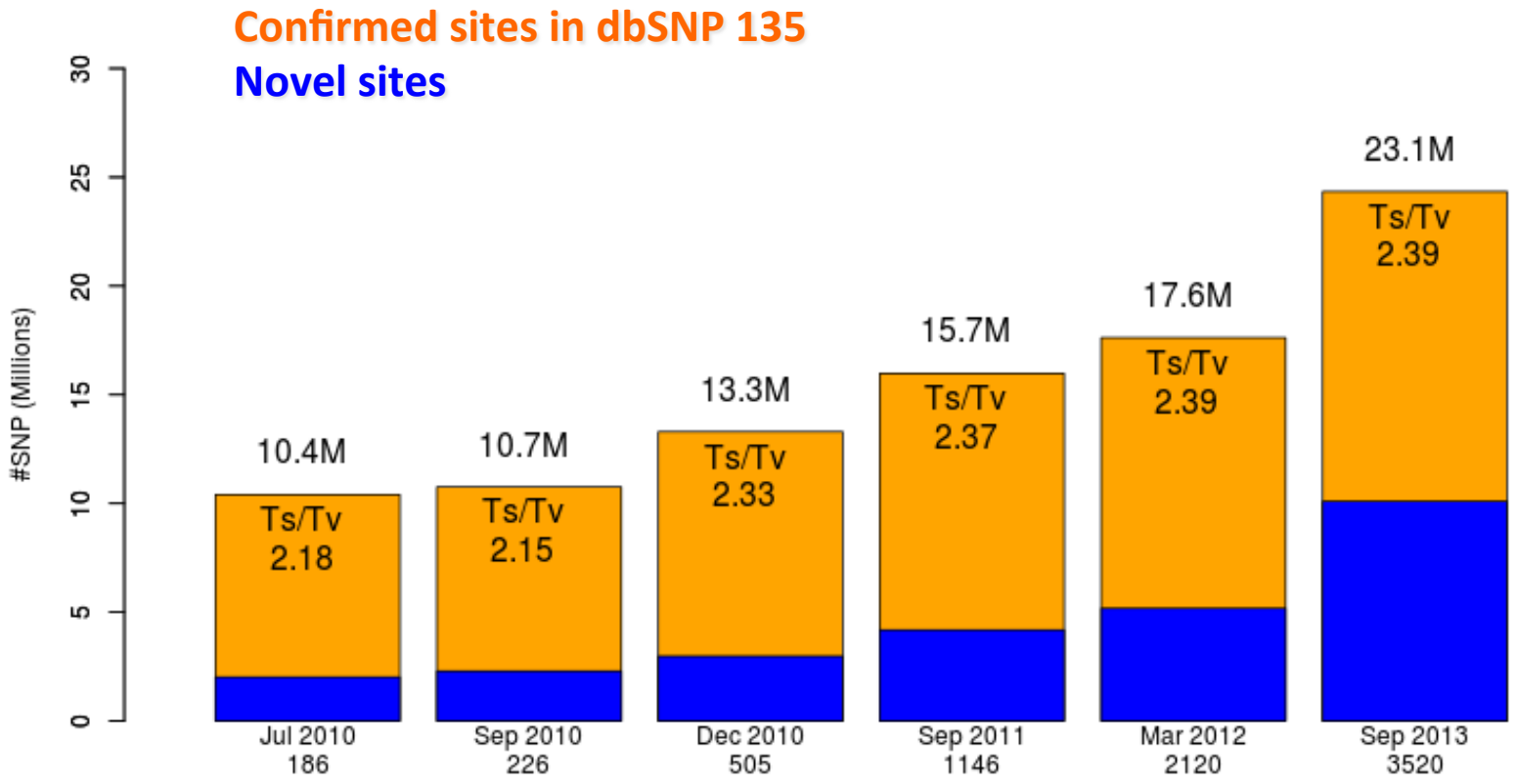
- **Deep extraction** of the genetic information

The Sardinian DNA Sequencing Project

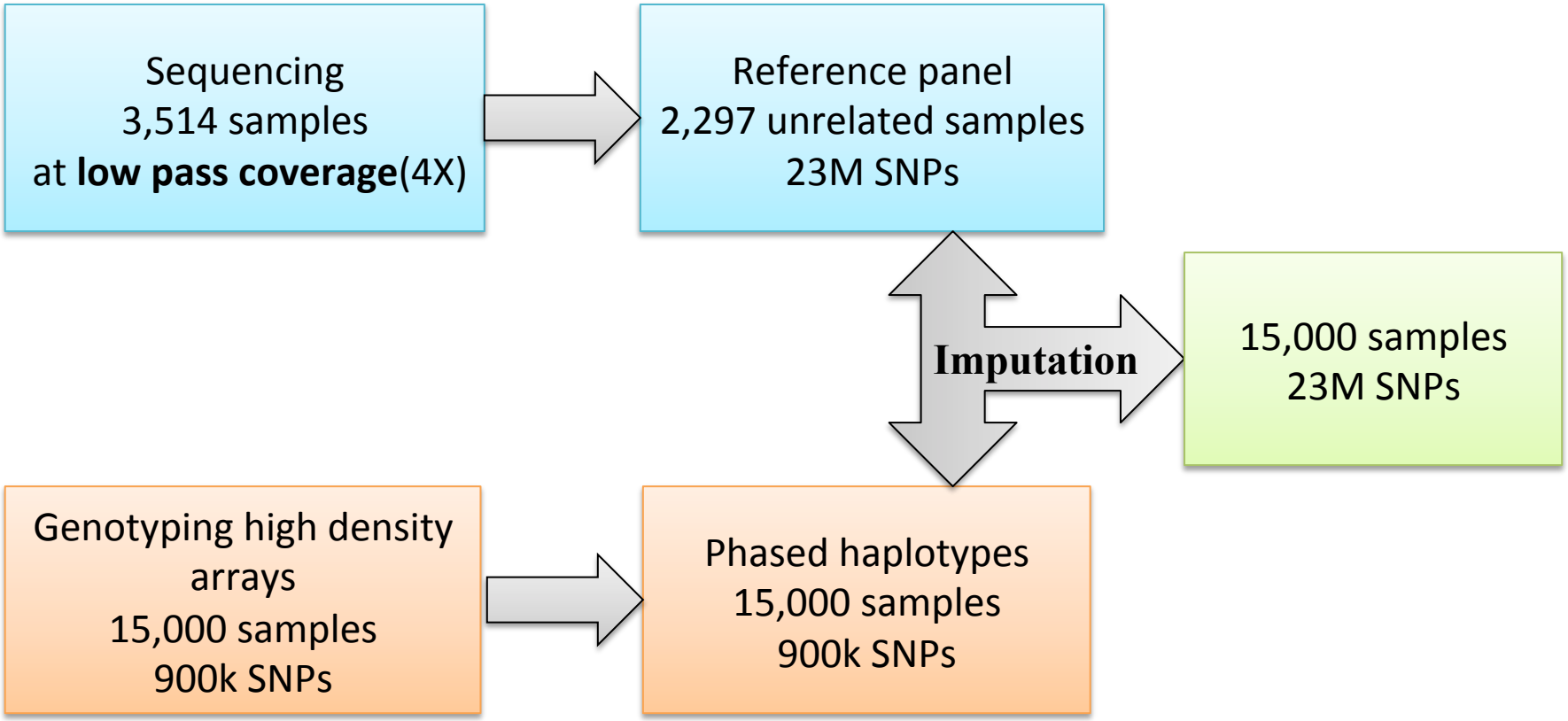
The sequencing project: goals

- To provide a more comprehensive representation of genetic variation in Sardinia
- To detect Sardinian founder variants not present in available GWAS arrays and in publicly available imputation panels
- To incorporate in one step both the detection and fine mapping experimental stages in sequencing based GWAS.

Milestones in 4 years:
More samples, more SNPs



Our low pass sequencing imputation strategy



nature genetics

Height-reducing variants and selection for short stature in Sardinia

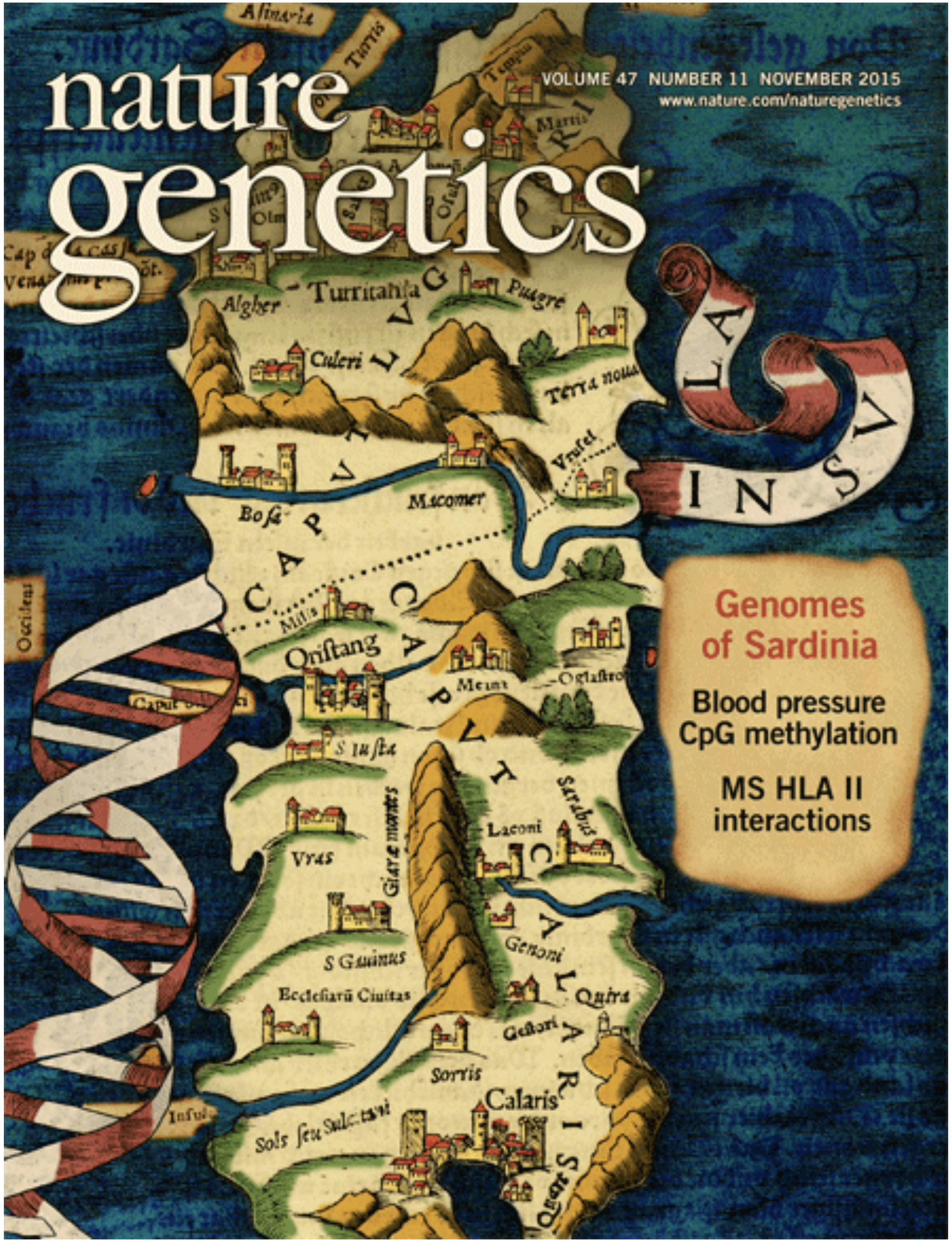
Magdalena Zoledziewska[1,19], Carlo Sidore[1,2,19], Charleston W K Chiang[3,19], Serena Sanna[1,19], Antonella Mulas[1,4], Maristella Steri[1], Fabio Busonero[1], Joseph H Marcus[5], Michele Marongiu[1], Andrea Maschio[1,2,6], Diego Ortega Del Vecchyo[7], Matteo Floris[1,4,8], Antonella Meloni[9], Alessandro Delitala[10], Maria Pina Concas[1], Federico Murgia[1], Ginevra Biino[11], Simona Vaccargiu[1], Ramaiah Nagaraja[12], Kirk E Lohmueller[3], UK10K Consortium[13], Nicholas J Timpson[14], Nicole Soranzo[15,16], Ioanna Tachmazidou[15], George Dedoussis[17], Eleftheria Zeggini[15], The Understanding Society Scientific Group[13], Sergio Uzzau[4,18], Chris Jones[8], Robert Lyons[6], Andrea Angius[1,8], Gonçalo R Abecasis[2,20], John Novembre[5,20], David Schlessinger[12,20] & Francesco Cucca[1,4,20]

Genome-wide association analyses based on whole-genome sequencing in Sardinia provide insights into regulation of hemoglobin levels

Fabrice Danjou[1,13], Magdalena Zoledziewska[1,13], Carlo Sidore[1–3], Maristella Steri[1], Fabio Busonero[1,2,4], Andrea Maschio[1,2,4], Antonella Mulas[1,3], Lucia Perseu[1], Susanna Barella[5], Eleonora Porcu[1–3], Giorgio Pistis[1–3], Maristella Pitzalis[1], Mauro Pala[1], Stephan Menzel[6], Sarah Metrustry[7], Timothy D Spector[7], Lidia Leoni[8], Andrea Angius[1,8], Manuela Uda[1], Paolo Moi[5,9], Swee Lay Thein[6,10], Renzo Galanello[5,9,12], Gonçalo R Abecasis[2,14], David Schlessinger[11,14], Serena Sanna[1,14] & Francesco Cucca[1,3,14]

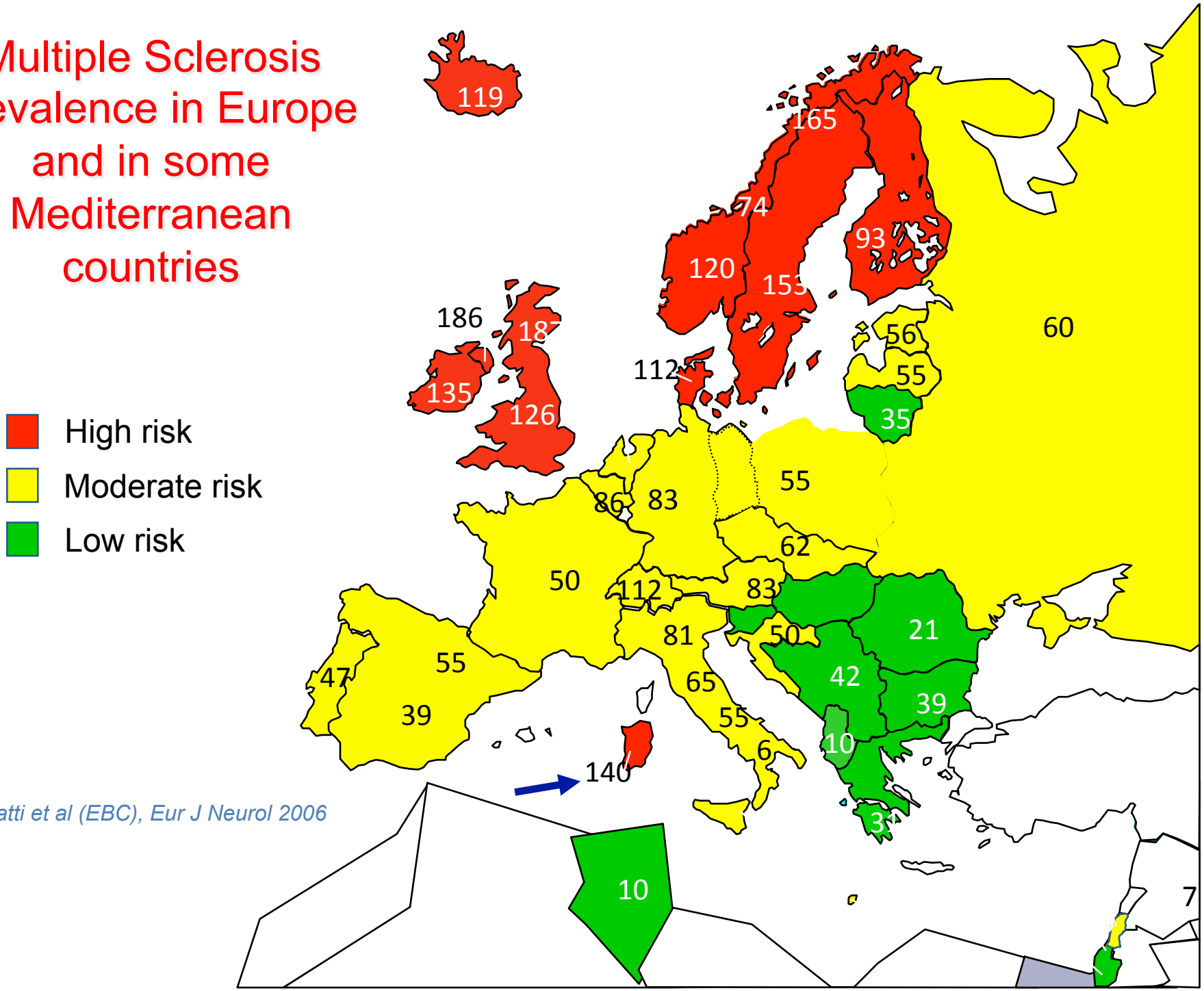
Genome sequencing elucidates Sardinian genetic architecture and augments association analyses for lipid and blood inflammatory markers

Carlo Sidore[1–3,19], Fabio Busonero[1,2,4,19], Andrea Maschio[1,2,4,19], Eleonora Porcu[1–3,19], Silvia Naitza[1,19], Magdalena Zoledziewska[1], Antonella Mulas[1,3], Giorgio Pistis[1–3], Maristella Steri[1], Fabrice Danjou[1], Alan Kwong[2], Vicente Diego Ortega del Vecchyo[5], Charleston W K Chiang[6], Jennifer Bragg-Gresham[2], Maristella Pitzalis[1], Ramaiah Nagaraja[7], Brendan Tarrier[4], Christine Brennan[4], Sergio Uzzau[8], Christian Fuchsberger[2], Rossano Atzeni[9], Frederic Reinier[9], Riccardo Berutti[3,9], Jie Huang[10], Nicholas J Timpson[11], Daniela Toniolo[12], Paolo Gasparini[13,14], Giovanni Malerba[15], George Dedoussis[16], Eleftheria Zeggini[10], Nicole Soranzo[10,17], Chris Jones[9], Robert Lyons[4], Andrea Angius[1,9], Hyun M Kang[2], John Novembre[18], Serena Sanna[1,20], David Schlessinger[7,20], Francesco Cucca[1,3,20] & Gonçalo R Abecasis[2,20]


nature genetics
VOLUME 47 NUMBER 11 NOVEMBER 2015
www.nature.com/naturegenetics
Genomes of Sardinia
Blood pressure
CpG methylation
MS HLA II interactions
INSVLA
Alghero
Turritania
Culeri
Terra noua
Macomer
Oristang
Laconi
Vras
Genoni
S Gauinus
Sorris
Calaris

The example of immune traits

Multiple Sclerosis prevalence in Europe and in some Mediterranean countries



Pugliatti et al (EBC), Eur J Neurol 2006

Goal:

- Our main goal is to use **genetics** as a tool to **dissect the biology of autoimmune diseases**, with a special emphasis on multiple sclerosis.

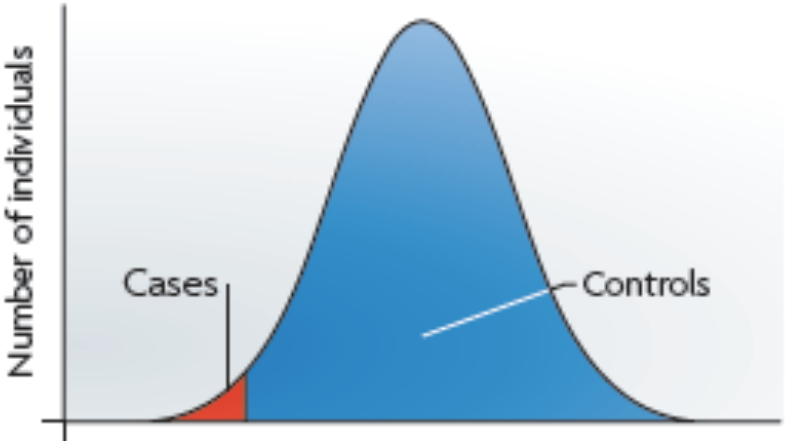
- Genetic associations with disease reveal **clues about mechanisms and pathways** that are typically much stronger and robust than can be had from epidemiological studies, and provide also **information about environmental exposures**.

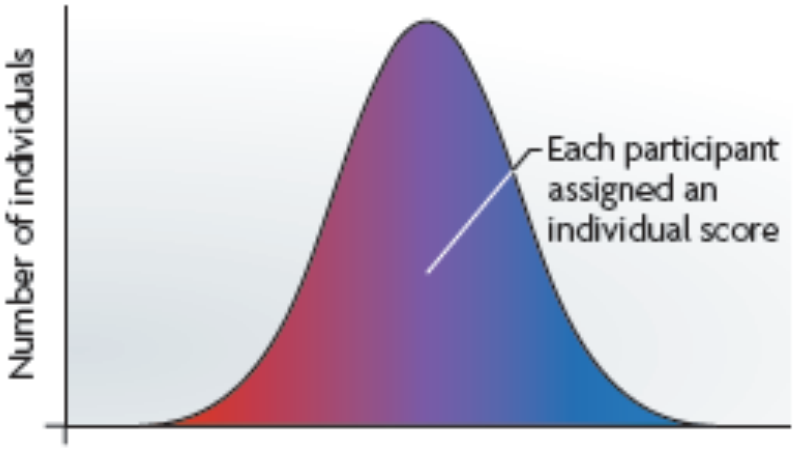
Our strategy:

- Searching for **coincident associations** between genetic variants controlling the circulating levels of **quantitative traits** (immune cell, antibody, cytokine blood levels) and affecting the risk for **diseases**.

- For example, this can reveal that variation in gene X modifies risk of disease Y by changing immune cell population Z.

- This approach is **distinct from hypothesis-driven comparisons** of cases and controls, which can be hampered by limited a priori knowledge and affected by second-order effects

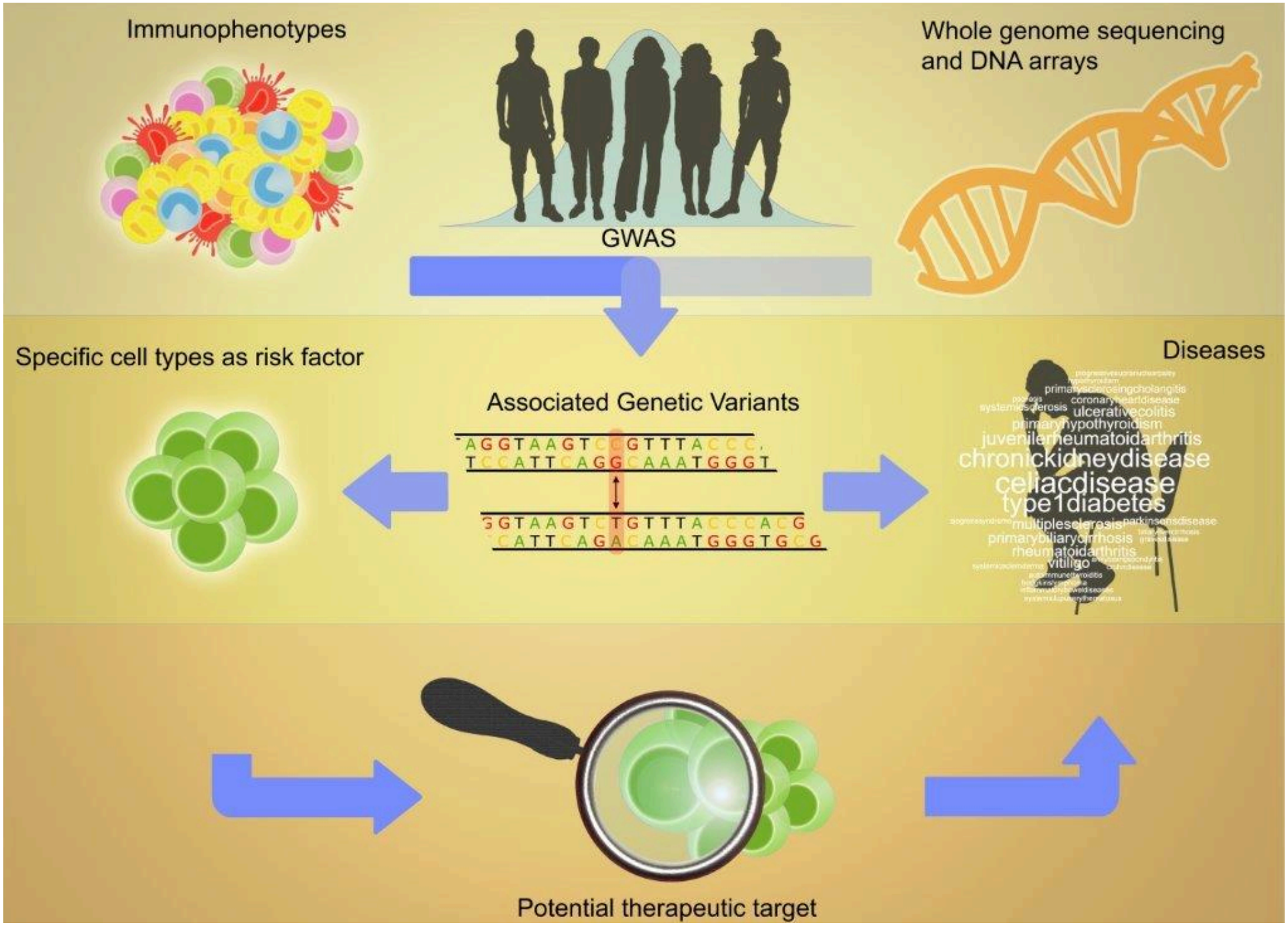




Qualitative trait

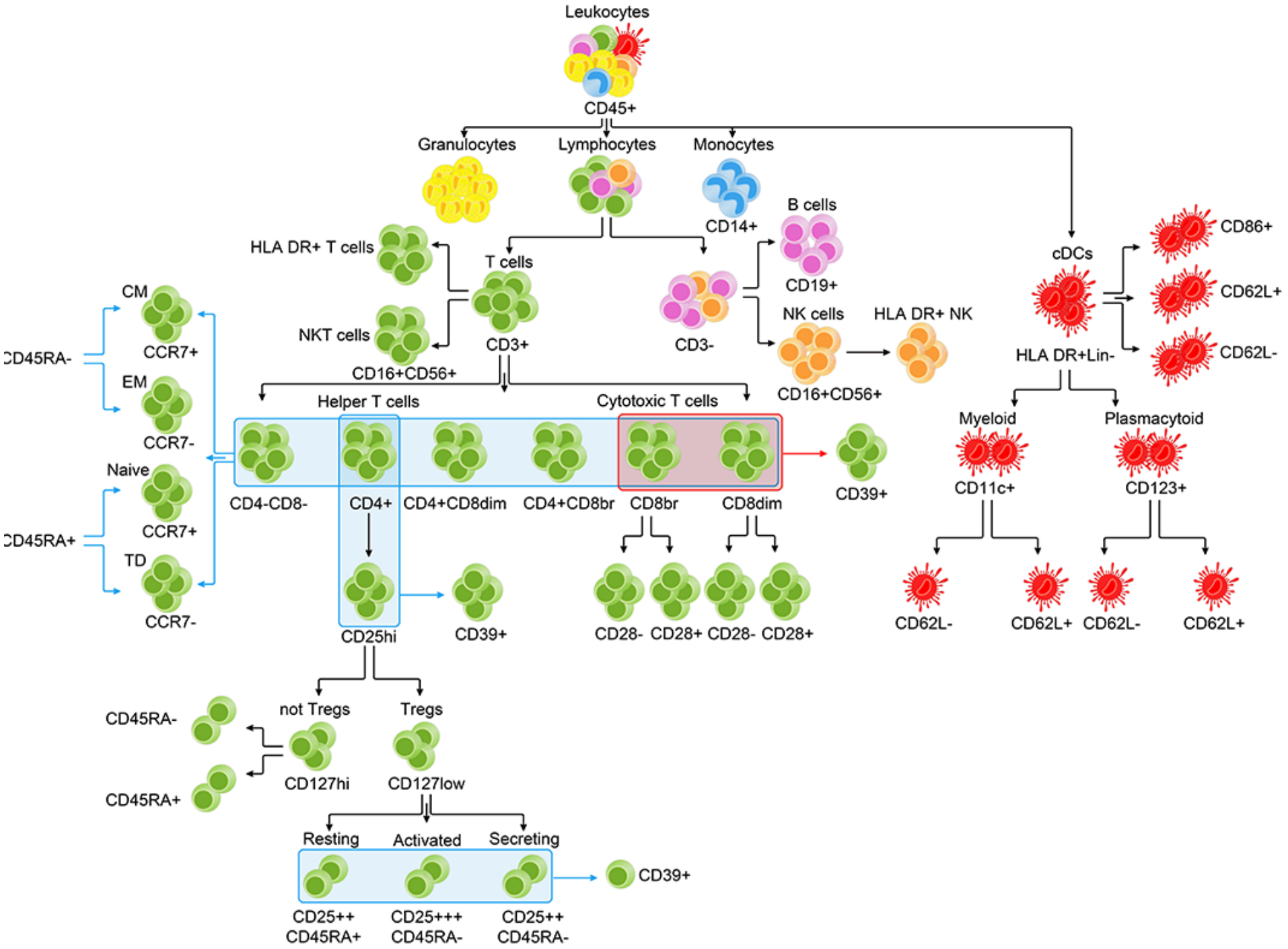
Quantitative trait

The Immune cell Project analysis based on fractionation of immune cell subtypes by FACS

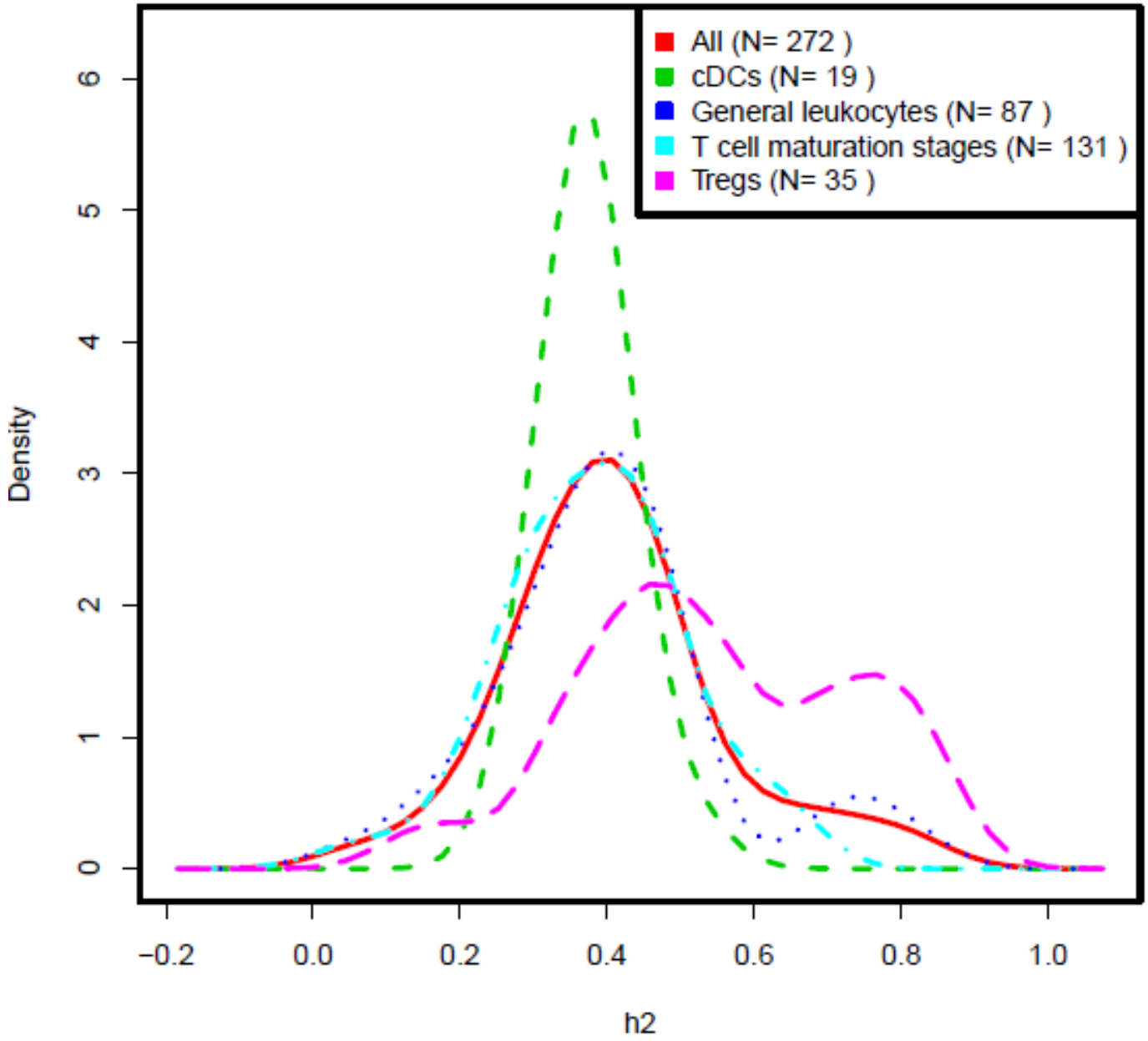

Immunophenotypes
GWAS
Whole genome sequencing
and DNA arrays
Specific cell types as risk factor
Associated Genetic Variants
AGGTAAGTCCGTTTACCC
TCCATTCAGGCAAATGGGT
GGTAAGTCTGTTTACCCACG
CATTCAGACAAATGGGTGCG
Diseases
primarysclerosingcholangitis
coronaryheartdisease
systemicsclerosis
ulcerativecolitis
primaryhypothyroidism
juvenilerheumatoidarthritis
chronickidneydisease
celiacdisease
type1diabetes
multiplesclerosis
parkinsonsdisease
primarybiliarycirrhosis
rheumatoidarthritis
vitiligo
Potential therapeutic target

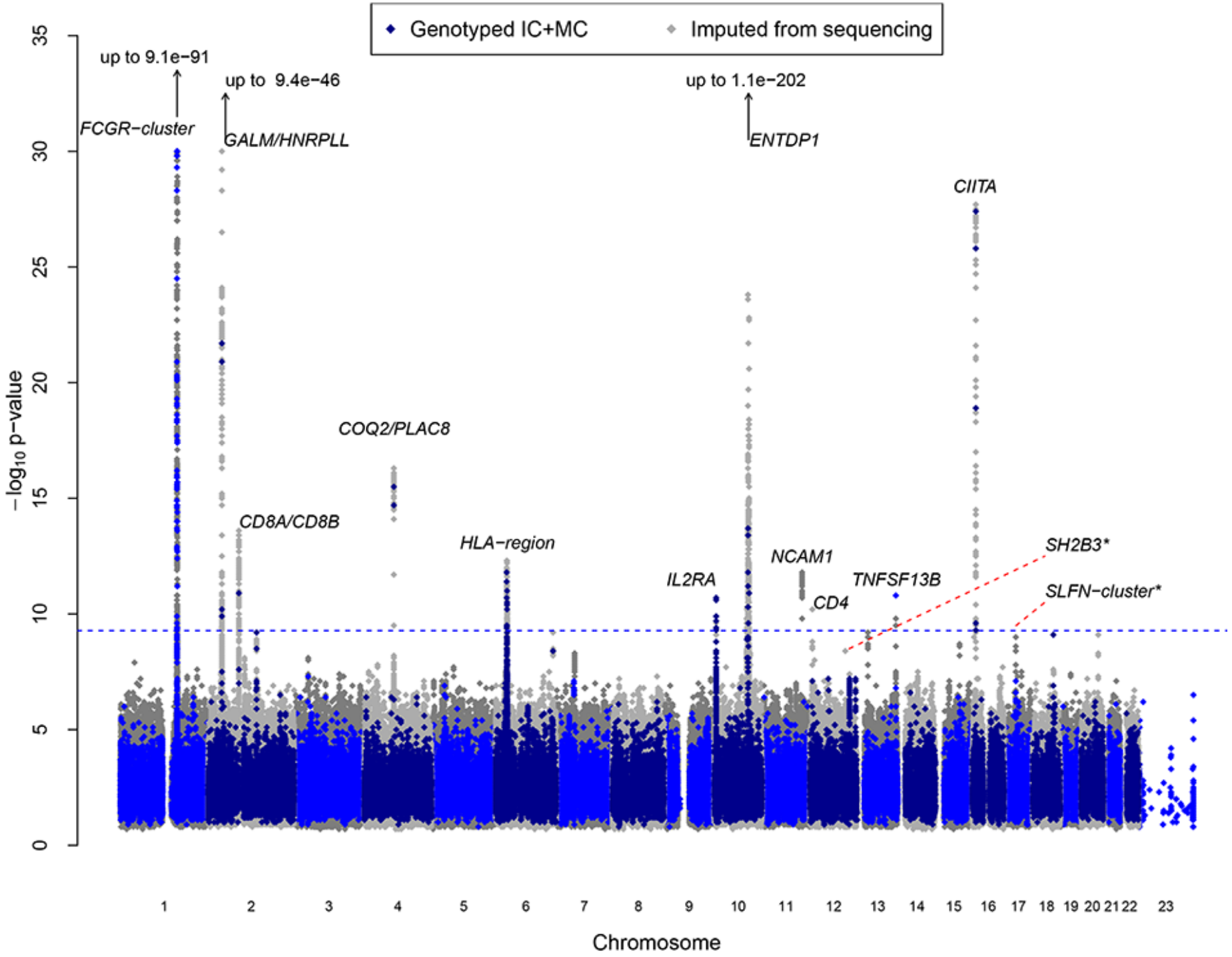
What have we measured in our first attempt?

- We measured **95 cells types**, and considered absolute cell counts, but also percentage respect to parental and grandparental lineages.
- A **total of 272 quantitative traits** were available for analyses, broadly divided in the following classes:
  - 87 General leukocyte sub-populations (B and T cells, Natural Killer, Monocytes and so on)
  - 35 T-regulatory cells
  - 19 Dendritic cells
  - 131 T cells maturation stages

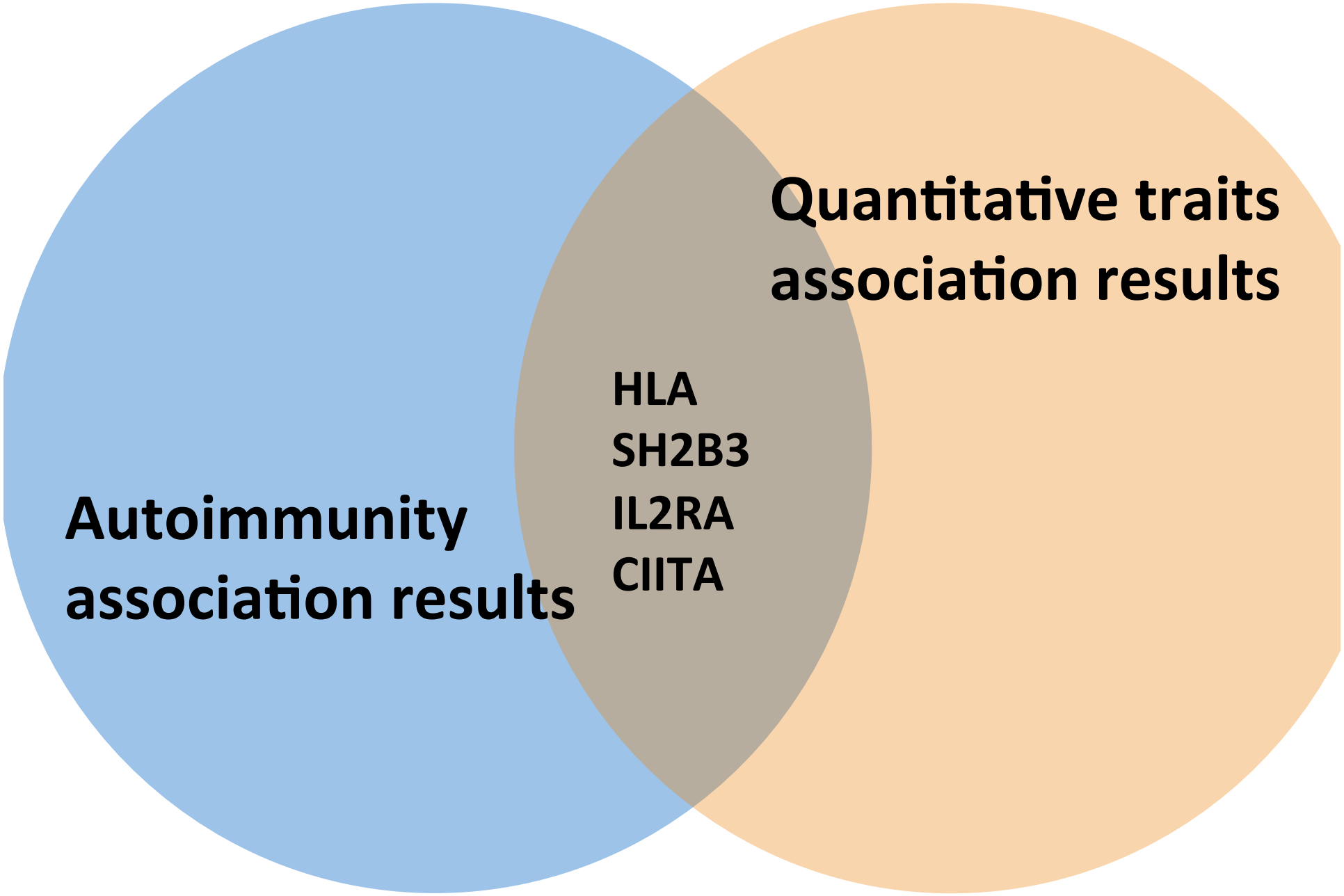

Leukocytes
CD45+
Granulocytes
Lymphocytes
Monocytes
CD14+
B cells
CD19+
HLA DR+ T cells
T cells
CD3+
NKT cells
CD16+CD56+
CD3-
NK cells
CD16+CD56+
HLA DR+ NK
cDCs
CD86+
CD62L+
CD62L-
HLA DR+Lin-
Myeloid
Plasmacytoid
CD11c+
CD123+
CD62L-
CD62L+
CD62L-
CD62L+
CM
CCR7+
EM
CCR7-
CD45RA-
Naive
CCR7+
TD
CCR7-
CD45RA+
Helper T cells
Cytotoxic T cells
CD4-CD8-
CD4+
CD4+CD8dim
CD4+CD8br
CD8br
CD8dim
CD39+
CD25hi
CD39+
CD28-
CD28+
CD28-
CD28+
not Tregs
Tregs
CD45RA-
CD45RA+
CD127hi
CD127low
Resting
Activated
Secreting
CD39+
CD25++
CD45RA+
CD25+++
CD45RA-
CD25++
CD45RA-

Heritability




Genotyped IC+MC
Imputed from sequencing
up to 9.1e−91
FCGR−cluster
up to 9.4e−46
GALM/HNRPLL
up to 1.1e−202
ENTDP1
CIITA
COQ2/PLAC8
CD8A/CD8B
HLA−region
IL2RA
NCAM1
CD4
TNFSF13B
SH2B3*
SLFN−cluster*
−log10 p−value
0
5
10
15
20
25
30
35
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
Chromosome

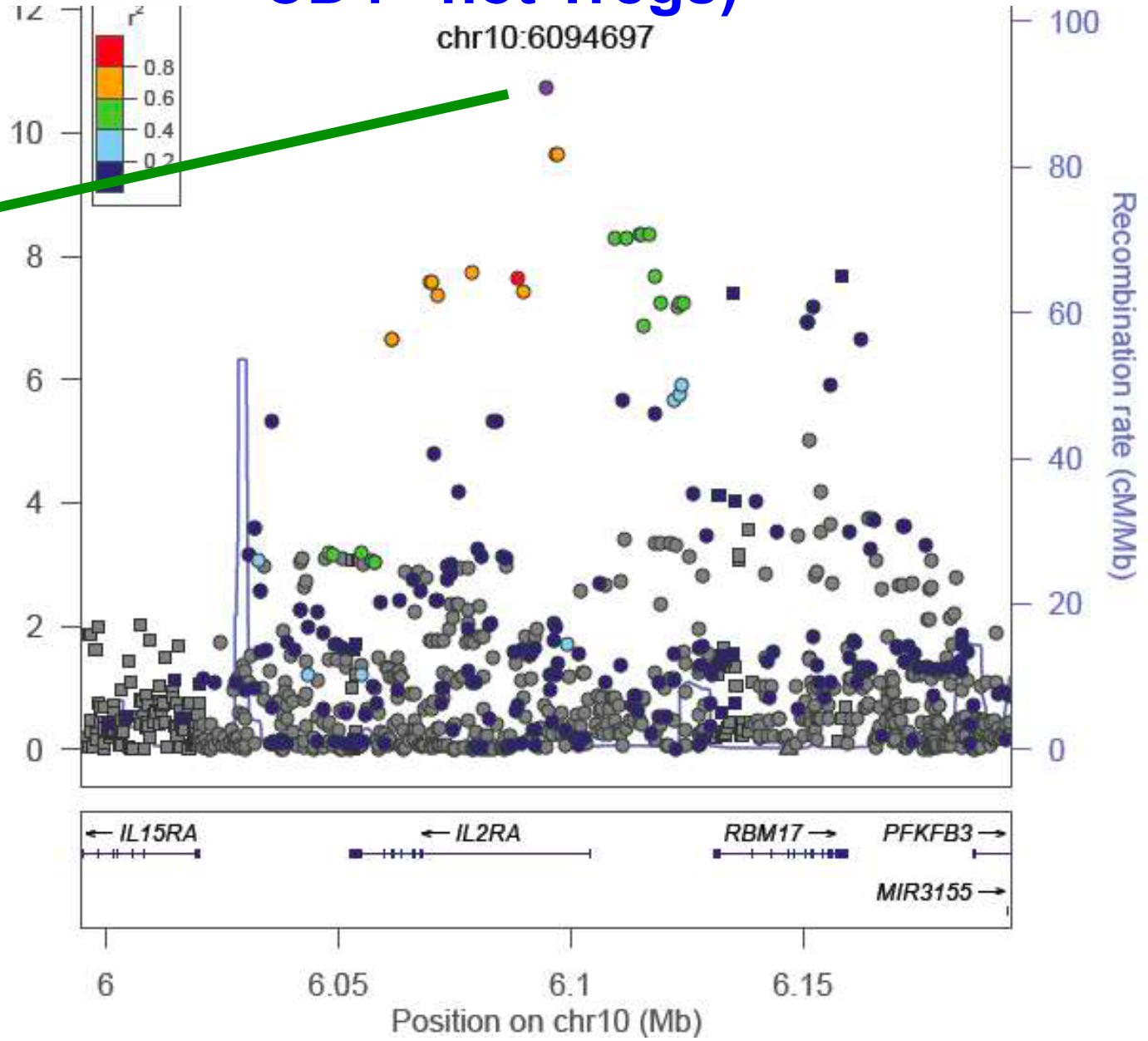
Are the same DNA variants associated with these quantitative traits also associated with immune-related diseases?


Autoimmunity
association results
HLA
SH2B3
IL2RA
CIITA
Quantitative traits
association results

An example:
a type 1 diabetes (T1D) locus correlated with a specific cell type

@IL2RA gene association with (CD45RA- CD25hi CD4+ not Tregs)

Protective for Type 1 diabetes



Cell

Resource

Genetic Variants Regulating Immune Cell Levels in Health and Disease

Valeria Orrù,[1,12] Maristella Steri,[1,12] Gabriella Sole,[1] Carlo Sidore,[1,2,3] Francesca Virdis,[1] Mariano Dei,[1] Sandra Lai,[1] Magdalena Zoledziewska,[1] Fabio Busonero,[1] Antonella Mulas,[1,3] Matteo Floris,[4] Wieslawa I. Mentzen,[1] Silvana A.M. Urru,[4] Stefania Olla,[1] Michele Marongiu,[1] Maria G. Piras,[1] Monia Lobina,[1,3] Andrea Maschio,[1,2] Maristella Pitzalis,[1] Maria F. Urru,[4] Marco Marcelli,[4] Roberto Cusano,[1,4] Francesca Deidda,[1,4] Valentina Serra,[1,3] Manuela Oppo,[4] Rosella Pilu,[1,4] Frederic Reinier,[4] Riccardo Berutti,[3,4] Luca Pireddu,[4,5] Ilenia Zara,[4] Eleonora Porcu,[1,3] Alan Kwong,[2] Christine Brennan,[11] Brendan Tarrier,[11] Robert Lyons,[11] Hyun M. Kang,[2] Sergio Uzzau,[3,6] Rossano Atzeni,[4] Maria Valentini,[4] Davide Firinu,[7] Lidia Leoni,[4] Gianluca Rotta,[8] Silvia Naitza,[1] Andrea Angius,[1,4] Mauro Congia,[9] Michael B. Whalen,[1] Chris M. Jones,[4] David Schlessinger,[10] Gonçalo R. Abecasis,[2] Edoardo Fiorillo,[1,12,*] Serena Sanna,[1,12,*] and Francesco Cucca[1,3,12,*]

[1]Istituto di Ricerca Genetica e Biomedica (IRGB), CNR, Monserrato 09042, Italy
[2]Center for Statistical Genetics, University of Michigan, Ann Arbor, MI 48109, USA
[3]Dipartimento di Scienze Biomediche, Università di Sassari, Sassari 07100, Italy
[4]CRS4, Parco Tecnologico della Sardegna, Pula, Cagliari 09010, Italy
[5]Università degli Studi di Cagliari, Cagliari 09010, Italy
[6]Laboratorio di Proteomica, Porto Conte Ricerche Srl, Tramariglio, Alghero 07041, Italy
[7]Dipartimento di Allergologia e Immunologia, Università di Cagliari, Cagliari 09124, Italy
[8]BD Biosciences Italia, Buccinasco, Milano 20090, Italy
[9]Dipartimento di Scienze Biomediche e Biotecnologie, Università di Cagliari, Cagliari 09124, Italy
[10]Laboratory of Genetics, NIA, Baltimore, MD 21224, USA
[11]University of Michigan Sequencing Core, University of Michigan Medical School, Ann Arbor, MI 48109, USA
[12]These authors contributed equally to this work
*Correspondence: edoardo.fiorillo@irgb.cnr.it (E.F.), serena.sanna@irgb.cnr.it (S.S.), francesco.cucca@irgb.cnr.it (F.C.)
http://dx.doi.org/10.1016/j.cell.2013.08.041

What are we measuring now?

• **47 new cells subtypes** (for a total of **119 quantitative traits** and **241 MFI**) in ~ **4,000 samples** (including ~ **300** subjects **>90 years old**):

- 22 B cells sub-populations (67 traits + 130 MFI)
- 8 Monocytes (19 traits + 52 MFI)
- 17 Myeloid Derived Suppressor Cells –MDSC (24 traits + 59 MFI)

Assessment **of >25 inflammation factors** and **soluble molecules** on the entire ProgeNIA cohort (~ **6,600 samples**):

• 7 Immunoglobulin levels (IgG and subtypes, IgM, IgA)
• ~20 cytokines and chemokines (including sBAFF, sCD25, HVEM)

How many immune variables overall?

- **142 cell types, 383 cellular traits** and **485 MFI**
- **25 inflammatory biomarkers/cytokines**

for a total of **~ 1,000 traits**

that we will refer to as the

“blood immunome”

for genetic and age and sex-related effects.

How this information impact ageing of the immune system (immunosenescence)?

ImmunoAgeing

Preliminary results:

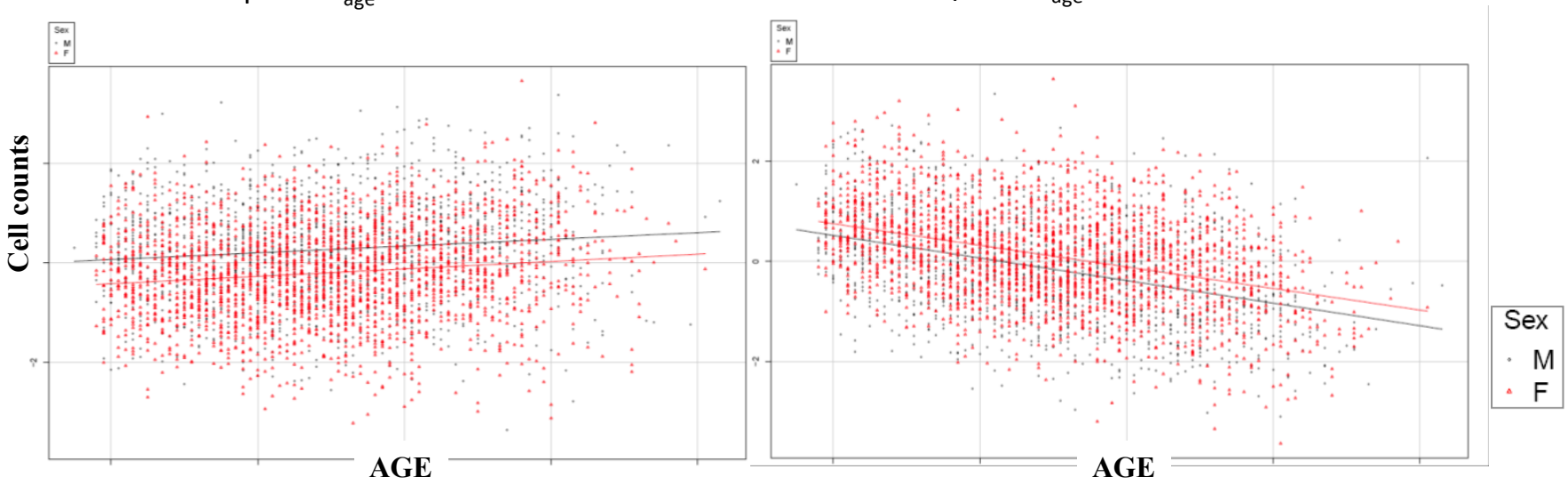
examples of age and sex related changes in specific immune cell numbers

NK

$p\text{-value}_{sex} = 1.63 \times 10^{-48}$
$p\text{-value}_{age} = 2.01 \times 10^{-07}$

CD45RA+ CD4+

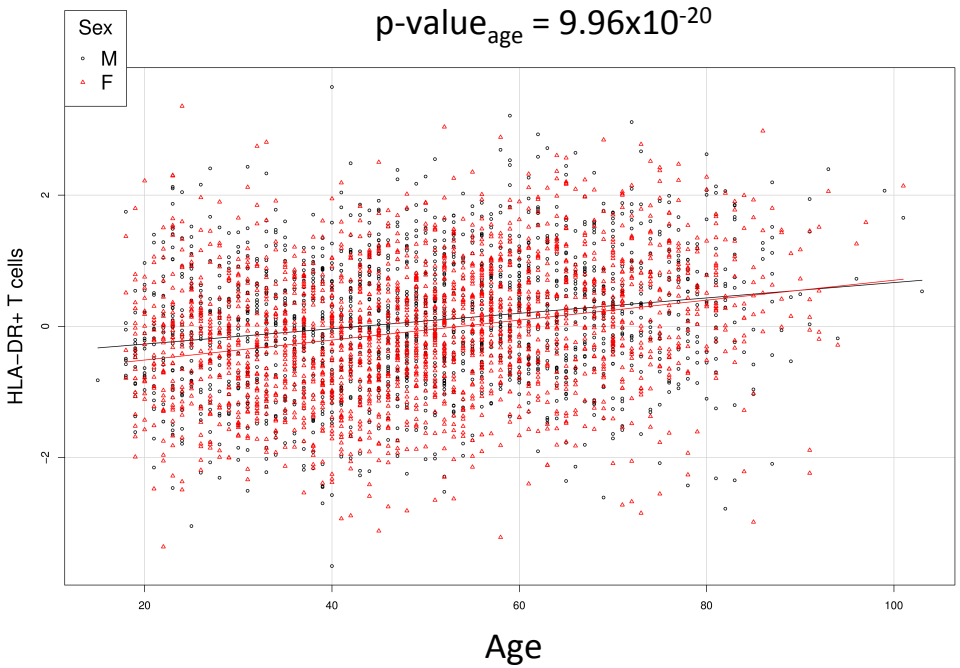
$p\text{-value}_{sex} = 3.11 \times 10^{-17}$
$p\text{-value}_{age} = 2.24 \times 10^{-60}$

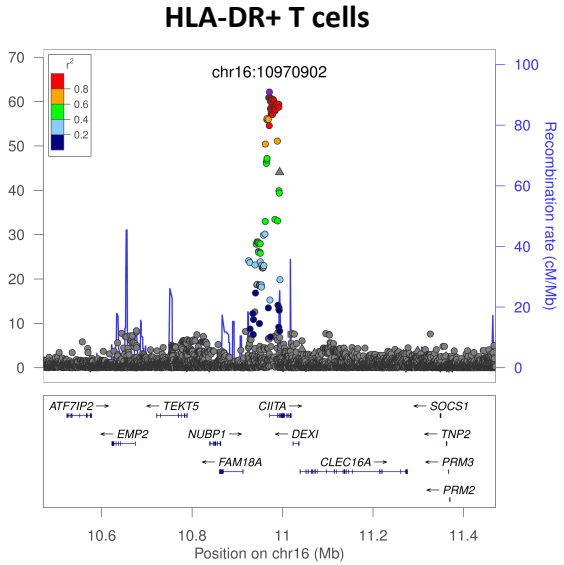


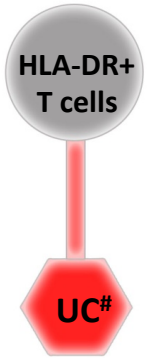
Statistically significant age changes and sex-related differences

ImmunoAgeing

Preliminary results:
traits affected by age and genetics







***CIITA* region chr. 16p13.13**

- Preliminary analyses on 3,500 ProgeNIA samples
- Genetic map including 23M SNPs and significance threshold: p-value=5.26x10^{-10}
- Samples stratified into four age intervals (18-40, 41-60, 61-80, and >80 yrs)
- Statistical significance threshold for ageing effects: p-value=5x10^{-4}

UC= Ulcerative Colitis

ImmunoAgeing

Conclusions

- Quantitative traits, represented by the levels of most of the immune cell populations exhibit substantial heritability

- We found a large number of genetic associations with these quantitative traits and establish their impact on immunosenescence

- Some of these variants coincide with previously and newly detected autoimmune diseases associations and reveal key players of immunosenescence.

- The data thus provide robust and unbiased hypotheses for biological studies.

Immunoageing Project

Francesco Cucca, Edoardo Fiorillo, Valeria Orrù, Maristella Steri **(IRGB-CNR)**

John Todd, Ken Smith, Eoin McKinney, Paul Lyons (**University of Cambridge)**

Beatrix Grubeck-Loebenstein (**Universitaet Innsbruck**)

Riccardo Bertini, Aldo Tagliabue (**ALTA**)