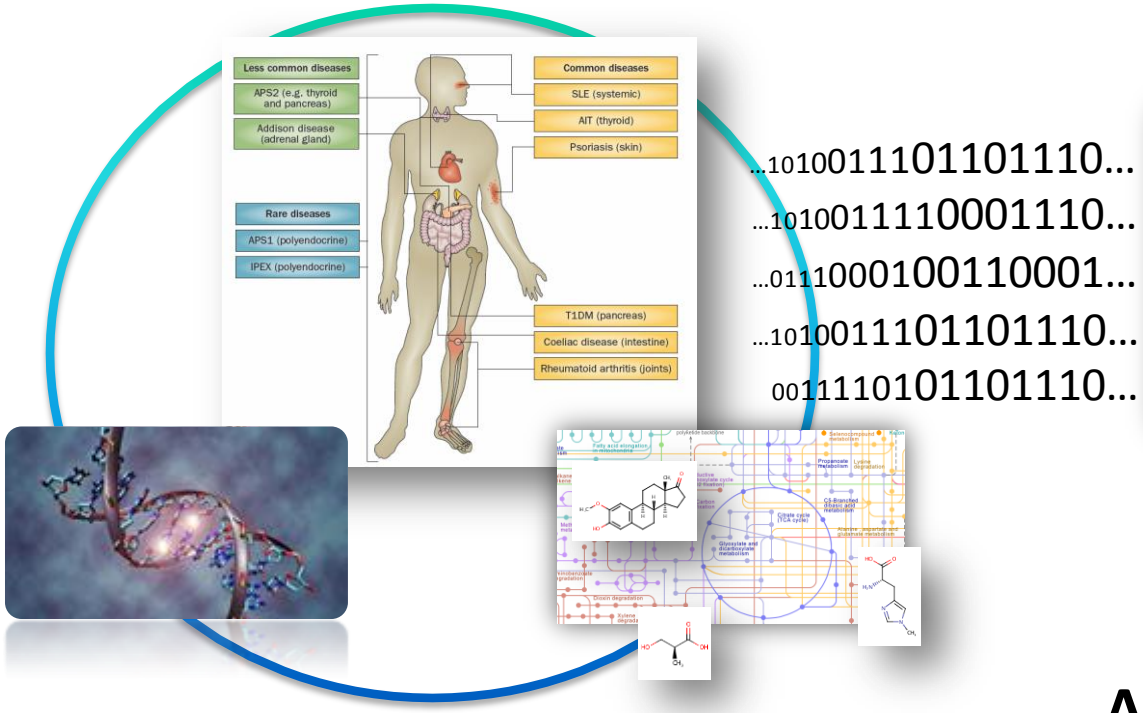
GENOMIC VARIATION ASSOCIATED TO THE METABOLOMICS OF AUTOIMMUNE DISEASES



...10100111101101110...
...10100111110001110...
...01110001001100001...
...10100111101101110...
...00111101011011110...

Antonio Julià, PhD
RHEUMATOLOGY RESEARCH GROUP
VALL HEBRON RESEARCH INSTITUTE
BARCELONA, SPAIN

9th RES Users' Conference
& 4th Annual HPC Advisory Council Spain Conference

WHAT ARE AUTOIMMUNE DISEASES?

Autoimmune diseases (AD) are a clinically diverse group of diseases that are caused by the inadequate activity of the immune system, reacting against the cells and tissues of our own body.

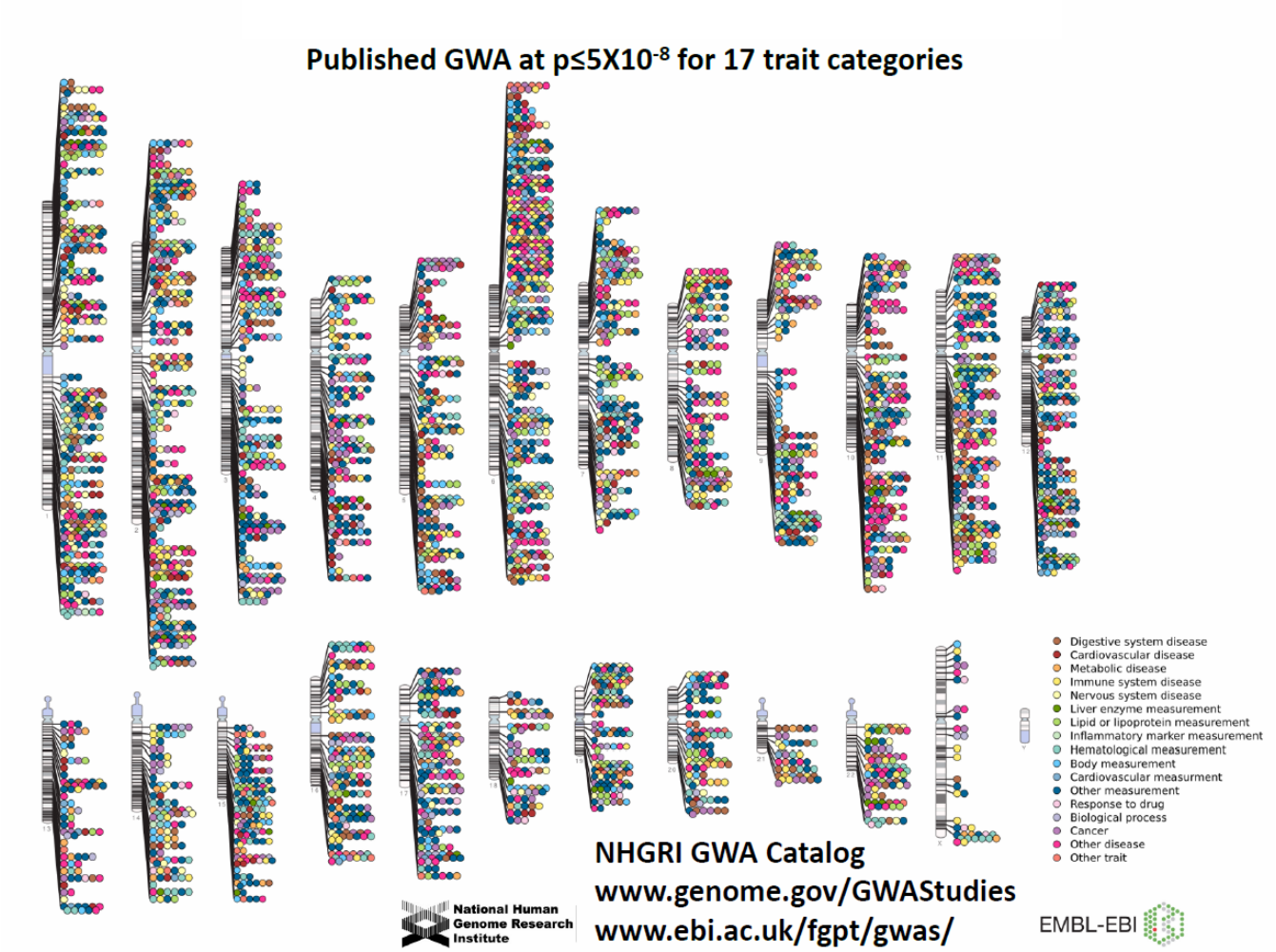
AUTOIMMUNE DISEASES ARE PREVALENT

- >5% of European population
- >100 types (many rare)
- 2nd highest cause of chronic illness
- Top cause of morbidity in women
- High health care costs (100.000m$ US vs 57.000 m$ cancer / 200.000m$ cardiov)

GENETICS IS A KEY DRIVER OF AUTOIMMUNE DISEASES

- Variation at the DNA level influences the risk to develop ADs
- Biotecnological breakthrough in genetics:
  - 2007-2014: microarrays
    - >500.000 Single Nucleotide Polymorphisms (SNPs) per individual
    - ~ 90% common variation covered
    - **Genome-Wide Association Studies (GWAS)**
  - 2014-onwards: next generation sequencing

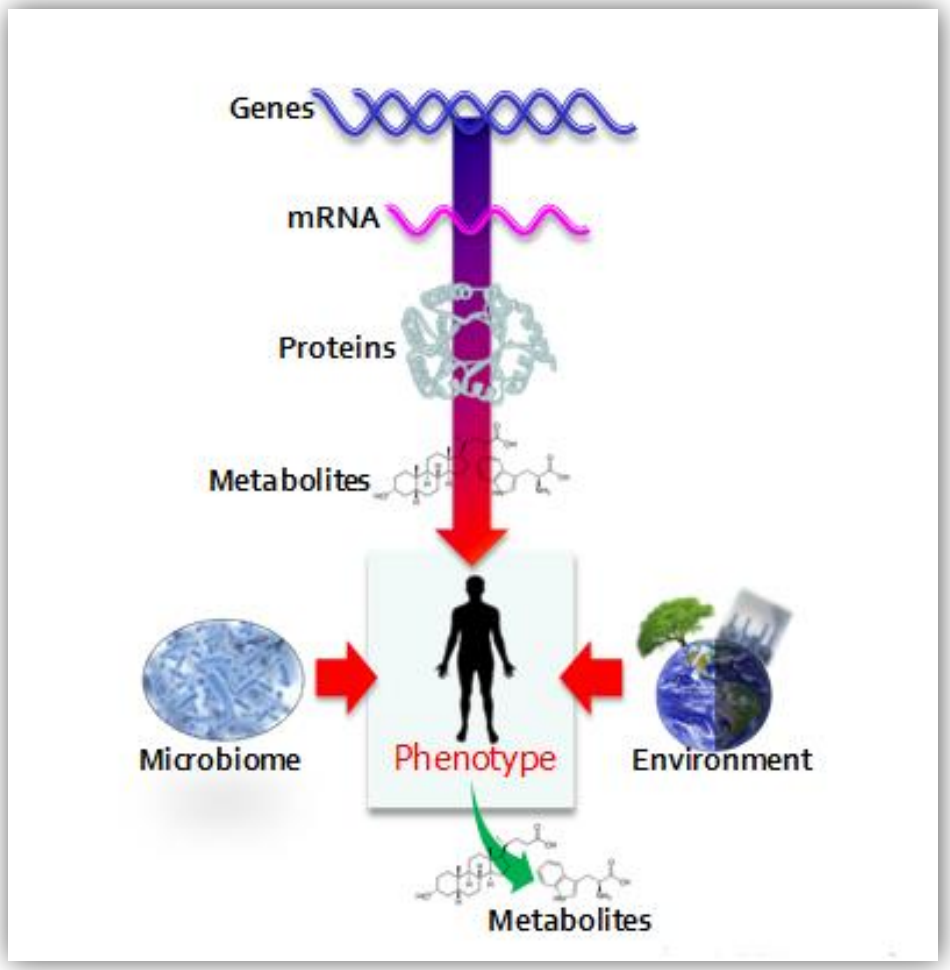
GWAS HAVE ALLOWED AN EXPONENTIAL DISCOVERY OF GENES ASSOCIATED WITH ADs



HETEROGENEITY IS A BOTTLENECK FOR THE STUDY OF ADs

- Most heritability in ADs still uncharacterized
- Like other complex diseases, ADs have a very high phenotypic variation
- Heterogeneity reduces our power to find relevant genetic variations and bio pathways
- Measuring individual variation at the molecular level (endophenotype), can increase our power to find new disease-relevant genes
- Metabolites: low molecular weight chemicals which are reactants or products of enzyme reactions

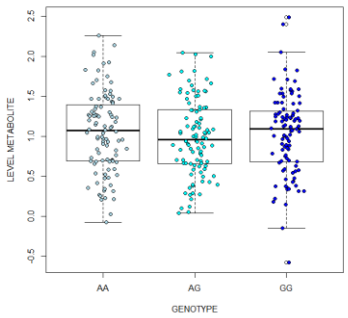
METABOLITES ARE CLOSER TO THE PHENOTYPE



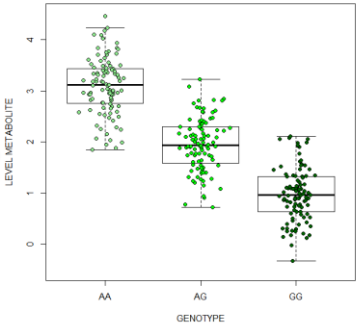
GENOTYPES ASSOCIATED TO METABOLITE LEVELS (METABOTYPES) CAN REVEAL NEW RELEVANT BIOLOGICAL PATHWAYS

- Link genetic variation with molecular variation within disease

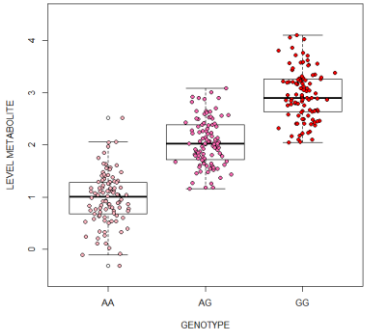
Normal



Disease X

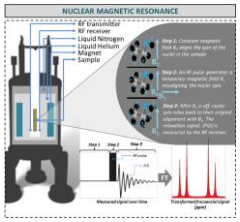


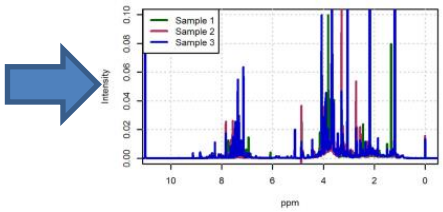
Disease Y



OBJECTIVE: IDENTIFY METABOTYPES IN ADs

- Target tissue: Urine
- Easy collection, ideal surrogate (blood)
- Largely unexplored so far
- Metabolic profile (metabolome) : Universitat Rovira i Virgili, Prof X Correig (CIBERDEM)
- Metabolome analysis using NMR
- NMR data analysis pipeline: FOCUS software



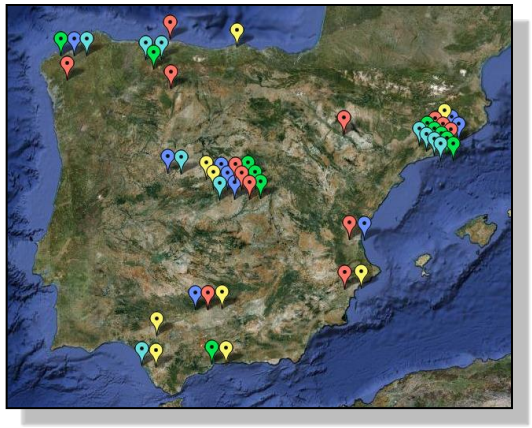


Alonso A et al *Anal Chem* '14

PERFORMING GWAS REQUIRES LARGE MEDICAL-BIOLOGIC SCIENCE CONSORTIUMS

IMID CONSORTIUM:

- >80 clinical departments from Spain
- Director Dr Sara Marsal (GRR)
- ~15,000 AD patients
- >1e6 biological samples
- One of largest repository AD
- Partially funded by Spanish grants (Singular & Strategic Proj., INNPACTO)



GOBIERNO DE ESPAÑA
MINISTERIO DE ECONOMÍA Y COMPETITIVIDAD

STUDY DESING INTEGRATES GENOMIC AND METABOLOMIC DATA

- Analysis of most prevalent ADs: Immune-Mediated Inflammatory Diseases:

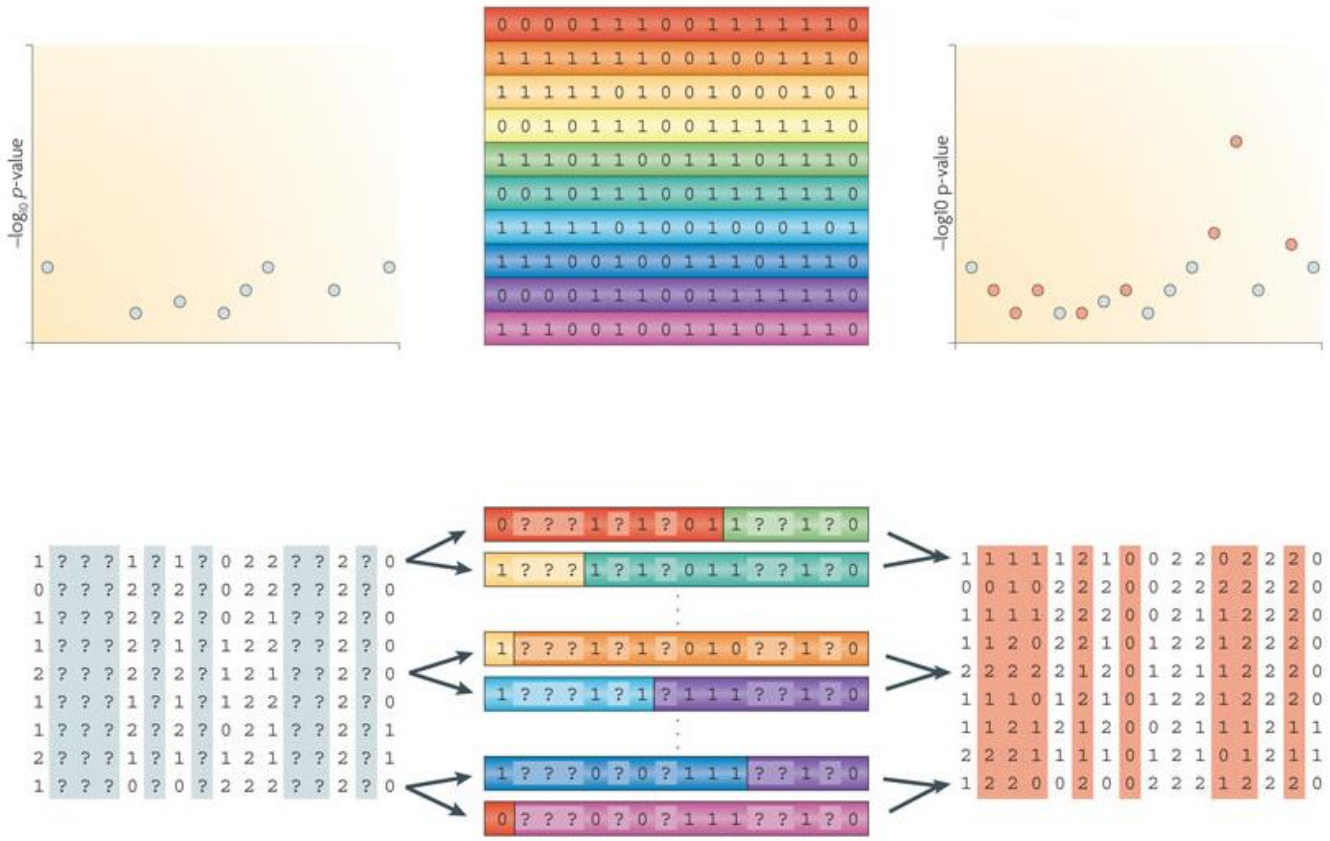
  - Rheumatoid Arthritis (RA)
  - Psoriasis (Ps)
  - Psoriatic Arthritis (PsA)
  - Crohn's Disease (CD)
  - Ulcerative Colitis (UC)
  - Systemic Lupus Erythematosus (SLE)

- Discovery cohort: 1,200 AD patients (RA, PsA, SLE, UC, CD, Ps)
- Validation cohort:1,200 AD patients (RA, PsA, SLE, UC, CD, Ps)
- Genotyping technology: Illumina Quad610microarrays (>550.000 SNPs)

CeGen
Centro Nacional de Genotipado

- Metabolite analysis: NMR of selected AD patients to minimze confunding (age, gender, etc.) and to represent different degrees of disease activity

IMPUTATION SIGNIFICANTLY INCREASES THE POWER OF GWAS

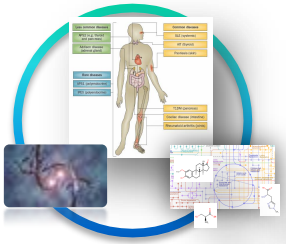


Marchini et al Nat Rev Genet '10

PRACE PROVIDES THE IDEAL FRAMEWORK FOR WHOLE GENOME IMPUTATION

- Objectives: set up analysis code for HPC for mGWAS
- **PRACE**: access to necessary high performance computing
- Preparatory type A – **Code Scalability testing**: summary, computer resources, simulation details, etc.
- Computational Genomics Group (BSC):
  - David Torrents
  - Josep M Mercader

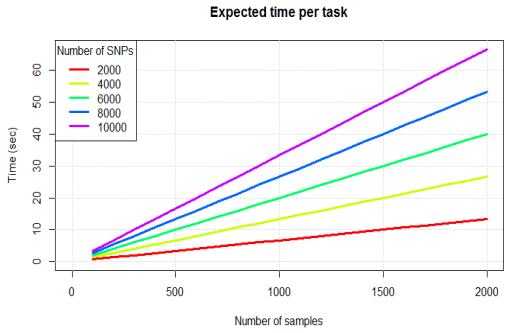
- Development and optimization of WG imputation & association testing pipeline (GWImp-COMPS)

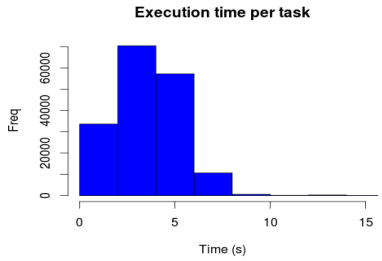


...10100111100011110...
...01110001001100001...
...10100111011011110...

PILOT STUDY SUCCESSFUL

- PRACE call for code scalability testing: 50,000 core hours in Mare Nostrum machine
- Chromosome 6:
  - ~6% genome
  - evaluate all code pipeline:
    i. SHAPEIT: haplotype phasing (O Delaneau Nat Meth '13)
    ii. IMPUTE: genotype imputation (B Howie Nat Genet'12)
    iii. SNPTEST: association testing (J Marchini Nat Genet '07)
  - scalability to all genome analysis
  - Greasy to parallelize executions
- 1000Genomes high density genetic data on 1,000 individuals (38e6 SNPs, 1.4e6 indels, 14K deletions) (Mc Vean Nature '12)





NEXT STEPS

- Application for regular PRACE project
- Pilot study allows:
  - Improvement of algorithms
  - accurate estimation of computational resources needed for large computational study
- Metabotypes associated with ADs with high statistical evidence:
  - Validation of mSNPs in independent sample cohort
  - This novel approach could identify new biological pathways that can lead to:
    - ✓New therapeutic targets
    - ✓Improved (early) diagnosis of ADs
    - ✓Better monitoring of disease

IMID Consortium

Jose Manuel Hernanz	Manuel Barreiro	Antonio Fernández-Nebro
Jose Luís Sánchez Carazo	Eugeni Domènech	Àlex Olivé
Francisco Vanadocha	Valle García	José Luis Andreu
Eduardo Fonseca	Esther García-Planella	José Luis Marenco de la Fuente
Lluís Puig	Ana Gutiérrez	MªAngeles Aguirre Zamorano
Carlos Ferrándiz	Juan Luís Mendoza	Paloma Vela
Esteban Dauden	Javier Pérez-Gisbert	Patricia Carreira
Emilia Fernández	Elena Ricart/ Julià Panés	Jose Javier Pérez-Venegas
David Moreno Ramírez	Cristina Saro	Ricardo Blanco / Víctor Martínez Taboada
José Luís López Estebaranz	Pilar Nos Mateu	Antonio Zea
Enrique Herrera	Fernando Gomollón García	Joan Miquel Nolla
Isabel Belinchón	Fernando Muñoz	Jose María Pego
Rafael Botella	Maribel Vera	Mercedes Freire
Ramon Pujol	Maria Esteve Comas	Elvira Díez
Juan Cañete	Jesús Rodríguez	Iñigo Jesús Rua-Figueroa Fernandez de Larrinoa
Jesús Tornero	José Javier Pérez Venegas	Alfredo Willisch Dominguez
Francisco Blanco	José Luís Fernández-Sueiro	César Díaz
Joan Maymo	Santiago Muñoz	Georgina Salvador
Javier Ballina	Santiago Muñoz	Alicia García
Benjamín Fernández	Raimon Sanmartí / Juan Cañete	José Antonio Mosquera
Àlex Olivé	Rubén Queiró	Santos Castañeda
Antonio Fernández Nebro	Jordi Gratacós	Simón Sánchez
Isidoro González	Carlos Montilla	Daniel Roig
Isabel Rotes	Pedro Zarco	Héctor Coromines
Carles Tomàs	Juan Carlos Torre-Alonso	Carlos González
Concha Delgado	Antonio Fernández Nebro	Pablo Unamuno

Dra Sara Marsal (Head of Group)
Raül Tortosa (IMID Biobank Coordinatior)
María López-Lasanta (IMID Clinical Coordinator)
Arnald Alonso
Nuria Palau
Adrià Aterido
Gabriela Ávila
Elena Granell
Carla Larroy
Pablo Pierrotti
Carolina Díaz
Andrea Pluma

David Torrents
Josep Mª Mercader
Silvia Bonàs
Friman Sánchez

Prof Xavier Correig
Marina Vinaixa
Miguel Ángel Rodríguez
Antoni Beltran