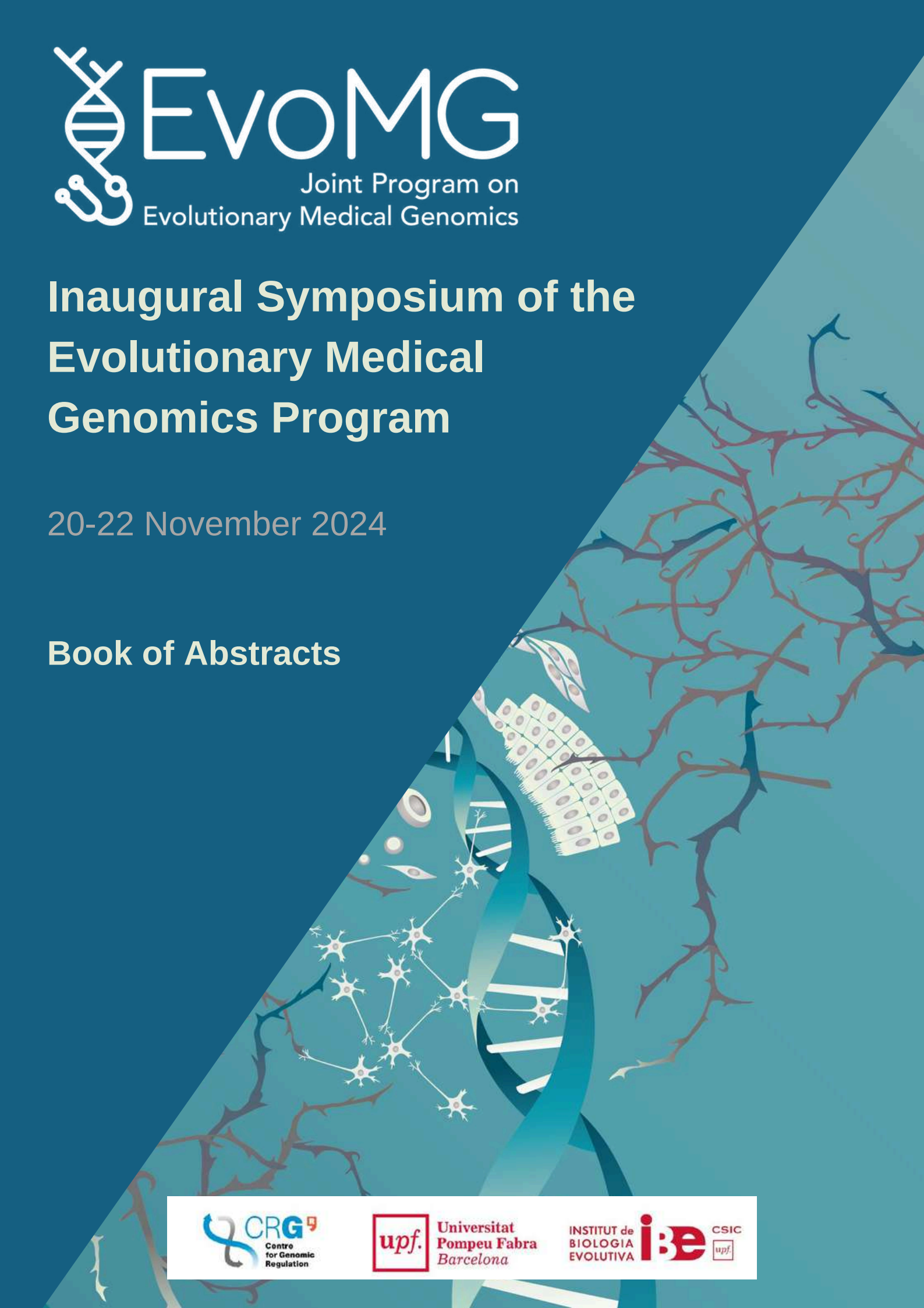
EvoMG

Joint Program on Evolutionary Medical Genomics

Inaugural Symposium of the Evolutionary Medical Genomics Program

20-22 November 2024

Book of Abstracts



CRG Centre for Genomic Regulation

upf. Universitat Pompeu Fabra Barcelona

INSTITUT de BIOLOGIA EVOLUTIVA iBE CSIC upf.

PROGRAM

DAY 1 - November 20th

13:30 Registration opens

14:30 Welcome and opening remarks

15:00 Keynoyte 1: **"Deep generative models for prediction and design in disease"**
Debbie Marks | Harvard Medical School
Chair: **Jonathan Frazer / Mafalda Dias,** EvoMG Program and CRG

Session I: Microbial pathogenicity and microbiome-host (co-)evolution
Chair: Amelie Baud, EvoMG Program and CRG

16:00 **"Within host evolution and transmission dynamics of the human microbiome"**
Mireia Valles-Colomer | EvoMG Program and University Pompeu Fabra

16:35 **"Tracing the origins and recent evolution of emerging yeast pathogens"**
Toni Gabaldon | Barcelona Supercomputing Centre-IRB Barcelona

17:10 Contributed talk: **"Machine learning classifiers reveal the evolutionary drivers of virulence and drug resistance in the fungal pathogen Candida parapsilosis"**
Miquel Àngel Schikora Tamarit | Barcelona Supercomputing Centre-IRB Barcelona

17:30 *Welcome drinks*

PROGRAM

DAY 2 - November 21st

Session II: Viral evolution, transposons and evolutionary immunology
Chair: Bernardo Rodriguez, EvoMG Program and CRG

09:30 **"How personalised is your immune repertoire?"**
Aleksandra Walczak | Ecole normale supérieure-PSL

10:05 **"How KRAB-zinc finger proteins emerge to fight invading retroviruses in mammals"**
Todd S. Macfarlan | Eunice Kennedy Shriver National Institute of Child Health and Human Development - NIH

10:40 Contributed talk: **"Adaptation in Human immune cells residing in tissues at the frontline of infections"**
Irepan Salvador | Centro Nacional de Análisis Genómico

11:00 *Coffee break*

Session III: Evolutionary history of disease
Chair: Elena Bosch, Evo MG Program and IBE (CSIC-UPF)

11:30 **"My ancestor was a Neanderthal. Should I see a doctor?"**
Tony Capra | UC San Francisco

PROGRAM

DAY 2 - November 21st

12:05 **"Population genetics and disease - how the past shapes health today"**
Astrid Iversen | University of Oxford

12:40 Contributed talk: **"Genomic Insights into the Evolutionary Nature of Autism Spectrum Disorder"**
Ariadna Bada i Navarro | University of Barcelona

13:00 *Lunch*

Session IV: Comparative genomic models of disease
Chair: Arnau Sebe Pedros, EvoMG Program and CRG

14:00 **"Transmissible cancers: when cancer cells become infectious agents"**
Elizabeth Murchison | University of Cambridge

14:35 **"Metabolic Adaptation to Nutrient Limitation in Vertebrates"**
Nicolas Rohner | University of Münster

15:10 Contributed talk: **"Evolutionary reversion in tumorigenesis"**
Yosuke Nagahata | IBE (CSIC-UPF)

15:30 *Coffee break and posters*

PROGRAM

DAY 2 - November 21st

Session V: Evolutionary cancer genomics
Chair: Francesc Calafell, Evo MG Program and IBE (CSIC-UPF)

17:00 **"Inference of the mutational processes and the mode of growth governing tumor evolution"**
Donate Weghorn | EvoMG Program and Centre for Genomic Regulation

17:35 **"The dynamics of genetic and epigenetic alterations in the decades preceding blood cancer diagnosis "**
Jamie Blundell | University of Cambridge

18:10 Contributed talk: **"Evolution of the p53/MDM2 interface leads to gain of functions in development and tumour suppression"**
Konstantinos Karakostis | IMIM

PROGRAM

DAY 3 - November 22nd

Session VI: Microbial pathogenicity and microbiome-host (co-)evolution
Chair: Arcadi Navarro, EvoMG Program, CRG and IBE (CSIC-UPF)

09:30 "**Global diversity, recurrent evolution, and recent selection on amylase structural haplotypes in humans"**
Peter H. Sudmant | UC Berkeley

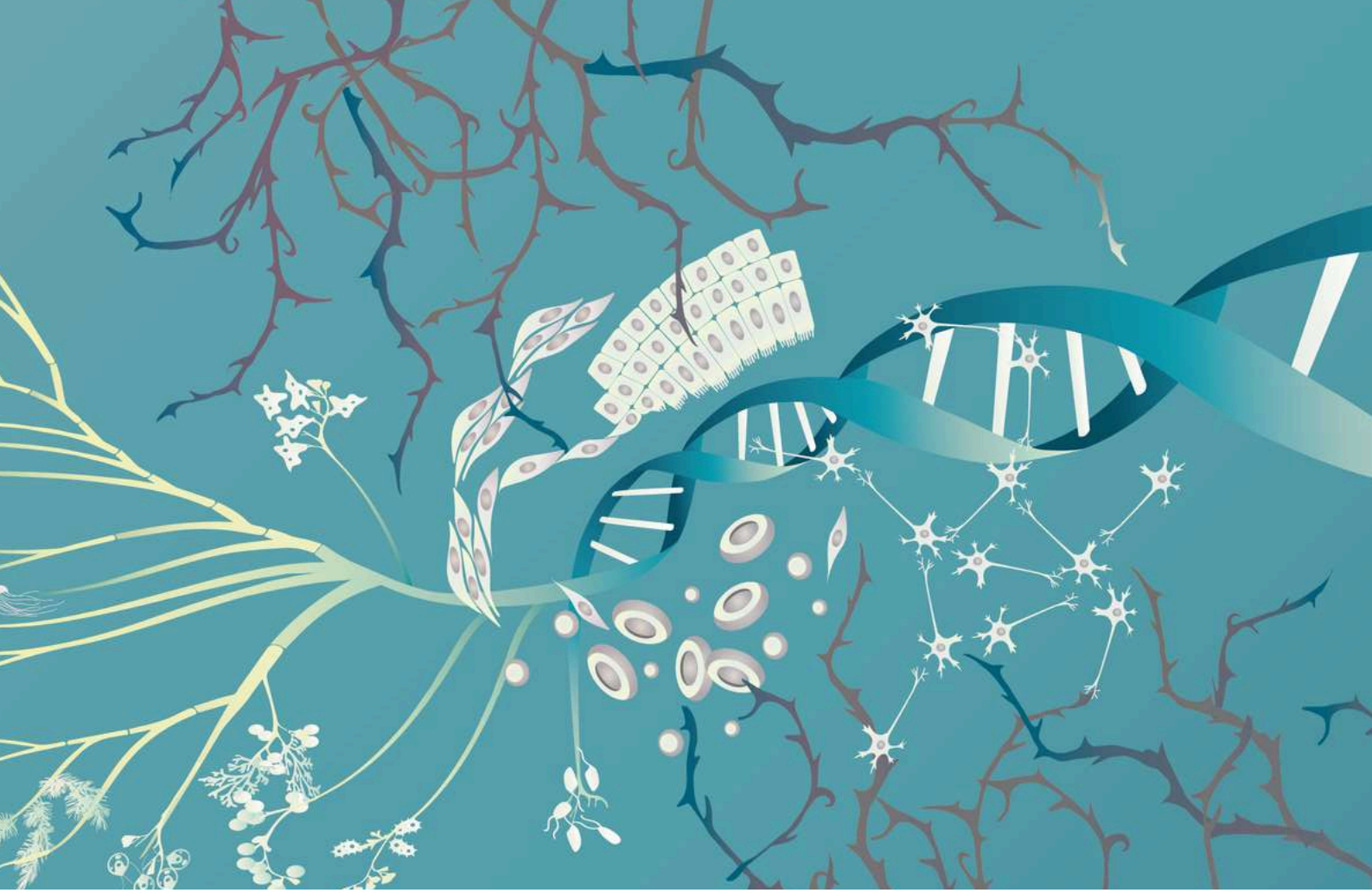
10:05 "**How do bats limit expected ageing and disease?"**
Emma C. Teeling | University College Dublin

10:40 Contributed talk: **"Interrogating osteoarthritis-related regulatory dynamics in primates using skeletal cell culture systems"**
Genevieve Housman | Max Planck Institute for Evolutionary Anthropology

11:00 Coffee break

11:30 Keynoyte 2: **"Evolutionary Therapy"**
Sandy Anderson | Moffitt Cancer Centre
Chair: **Manuel Irimia**, Coordinator of the EvoMG Program and UPF-CRG

12:30 Closing remarks



Keynote speakers

Sandy Anderson

Moffitt Cancer Centre

“Evolutionary Therapy”

Our current approach to cancer treatment has been largely driven by finding molecular targets, those patients fortunate enough to have a targetable mutation will receive a fixed treatment schedule designed to deliver the maximum tolerated dose (MTD). Cancers are complex evolving systems that adapt to therapeutic intervention through a suite of resistance mechanisms, therefore whilst MTD therapies generally achieve impressive short-term responses, they unfortunately give way to treatment resistance and tumor relapse. The importance of evolution during both tumor progression, metastasis and treatment response is becoming more widely accepted. However, MTD treatment strategies continue to dominate the precision oncology landscape. Here we discuss evolutionary therapy, a proactive therapeutic approach that changes and evolves with the tumor being treated. Due to the dynamic feedback between changing treatments and the evolving tumor, mathematical models are essential to drive treatment switch points and predict appropriate dosing and drug combinations. We will consider the importance of using treatment response as a critical driver of subsequent treatment decisions, rather than fixed MTD strategies that ignore it.

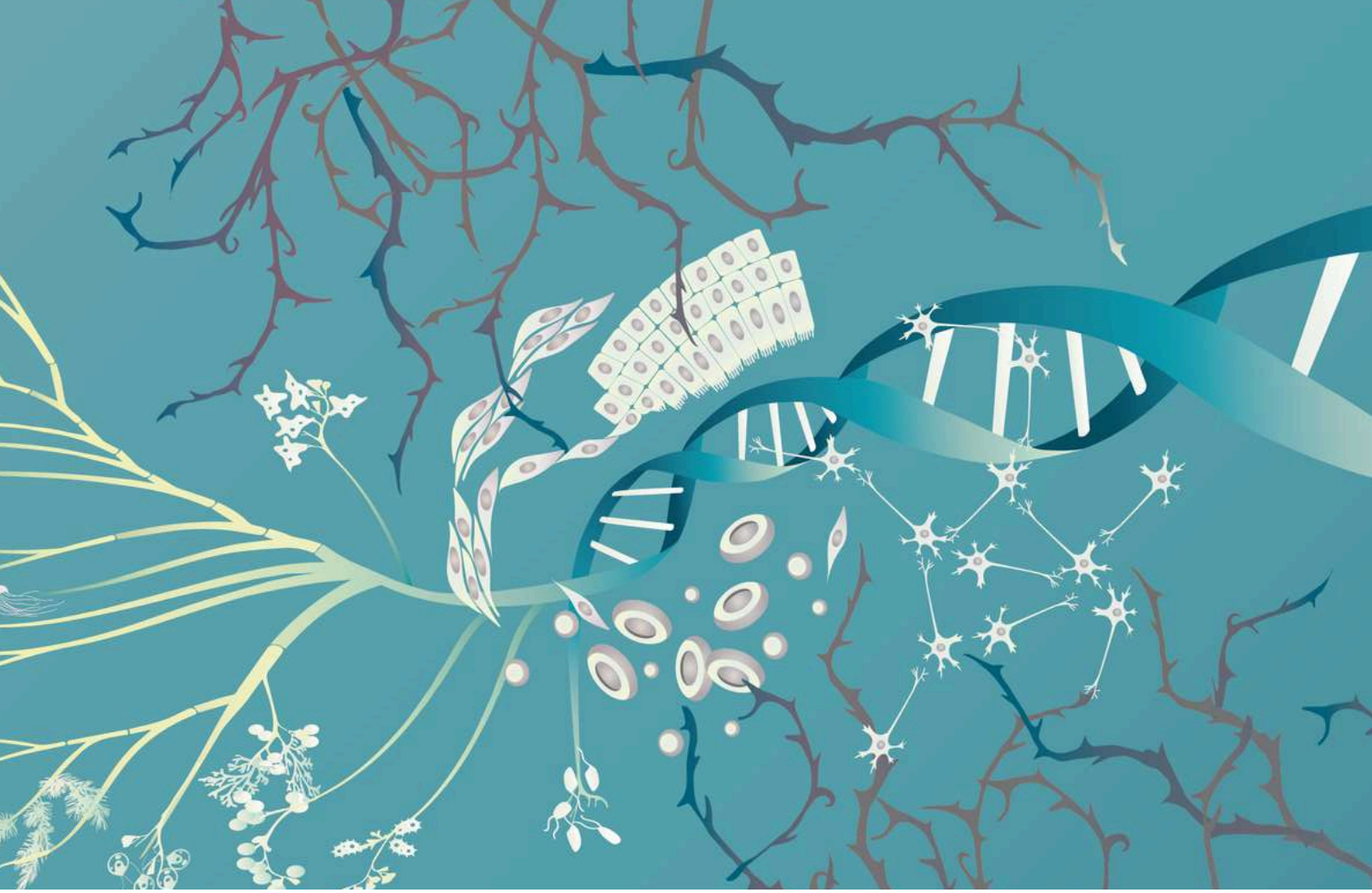
We will also consider using mathematical models to drive treatment decisions based on limited clinical data. Through the integrated application of mathematical and experimental models as well as clinical data we will illustrate that, evolutionary therapy can drive either tumor control or extinction. Our results strongly indicate that the future of precision medicine shouldn’t only be in the development of new drugs but rather in the smarter evolutionary, and model informed, application of preexisting ones.

Debbie Marks

Harvard Medical School

“Deep generative models for prediction and design in disease”

There’s now an amazing opportunity to accelerate discovery across important challenges by using computation tightly coupled to biological experiments and clinical medicine. I will describe some recent approaches from my lab for these challenges where we have developed new machine learning methods that can exploit the enormous natural sequence diversity and our ability to synthesize DNA at scale. The applications will include our recent work predicting the effects of human genetic variation on disease, anticipation of viral escape from the host immune system for vaccine design, protein design for enzyme optimization and smart antibody libraries. I will end by describing the many and important remaining challenges that I hope will inspire you to work on them!



Invited speakers

Tony Capra

UC San Francisco

“My ancestor was a Neanderthal. Should I see a doctor?”

In this talk, I will describe how ancient DNA, population-scale genetic databases, and machine learning are enabling us to understand how interactions between ancient groups influence phenotypes in modern populations and what they can teach us about uniquely human traits.

Astrid Iversen

University of Oxford

“Population genetics and disease - how the past shapes health today”

This presentation will explore our recent research on the genetic ancestry of modern Europeans and how the historical context of European populations influences their disease risk today, using multiple sclerosis (MS) as an example. Modern humans settled Europe in three main waves: hunter-gatherers arrived around 45,000 years ago (and spit into Western hunter-gatherers and Eastern hunter-gatherers), farmers from the Middle East about 11,000 years ago, and pastoralists from the steppes of western Asia and eastern Europe reached Europe approximately 5,000 years ago. Notably, dispersal patterns indicate that pastoralists primarily settled in northern and Northwestern Europe, while farmers spread through Europe in a South-to-North gradient from Anatolia (present-day Turkey/Türkiye), intermixing with hunter-gatherers to varying extents on the way. As a result, modern Europeans have different proportions of ancestry from these groups, which affects their susceptibility to various diseases.

Our findings suggest that the risk of MS, an autoimmune and debilitating disease characterized by an overactive immune response, is, to a large extent, linked to the genetic contributions of ancient pastoralists.

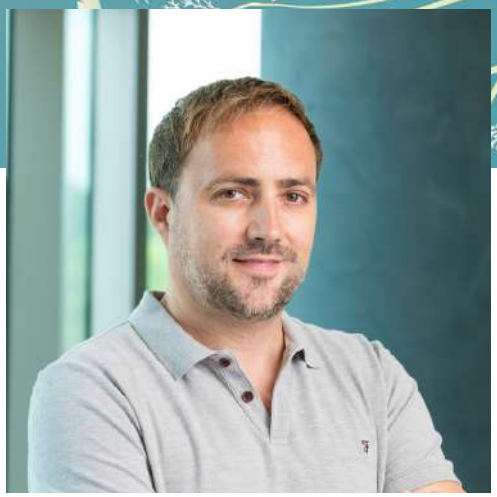
We found a higher proportion of pastoralist ancestry in northern Europeans than southern Europeans, shedding light on the north-south gradient of MS prevalence. We suggest these genetic variants might have provided advantages against novel, more complex and/or intense pathogen challenges linked to the transition from hunter-gatherer societies to farming and pastoral lifestyles. Future studies of ancient human and animal genomes from Europe and other parts of the world, including analyses of pathogens, are essential to elucidate the relationship between infectious diseases, genetic footprints of selection and current autoimmune risks.

Elizabeth Murchison

University of Cambridge

“Transmissible cancers: when cancer cells become infectious agents”

Cancer arises when mutations drive cells of the body to abandon their usual functions and to instead embark upon a “selfish” evolutionary programme underpinned by abnormal growth. Most cancers exist only within the bodies of the hosts that spawn them; rarely, however, cancers can acquire adaptations allowing them to spread between individuals. In such transmissible cancers the cancer cells themselves become agents of infection. Elizabeth Murchison will discuss recent research on the origins and evolution of the naturally occurring mammalian transmissible cancers affecting dogs and Tasmanian devils.



Nicolas Rohner

University of Münster

“Metabolic Adaptation to Nutrient Limitation in Vertebrates”

Adaptation to food deprivation is widespread among animal species, reflecting the intimate connection between genotype, phenotype, and the environment. However, the genetic basis of physiological adaptations to nutrient availability remains an unresolved challenge of both organismal biology and modern evolutionary genetics. We are using the cavefish Astyanax mexicanus as a promising research organism to unravel the genetic basis of starvation resistance. A. mexicanus exists in two forms: a river-dwelling surface fish and a blind, depigmented cavefish. Whereas the surface forms live in a rich ecological environment, multiple distinct cave populations have evolved metabolic adaptations to nutrient limitations in caves. Importantly, the surface and cave morphs remain interfertile and can be bred in the laboratory. Using recently developed genetic and genomic tools, we have shown that cavefish evolved a massive capacity for fat storage due to increased appetite, adipogenesis, and lipogenesis. In addition, we found that cavefish display elevated blood sugar levels and insulin resistance caused by a mutation in their insulin receptor.

Unlike humans with the same mutation, cavefish do not display diabetes markers and live long and healthy lives. Furthermore, cavefish develop hypertrophic visceral adipocytes without obvious signs of inflammation due to reduced amounts of pro-inflammatory cytokines. In a more recent series of studies, we showed that cavefish are thriftier due to decreased muscle mass, improved glycogen production, and efficient recycling of amino acids. As all these extreme adaptations have no negative consequences on the metabolic health, immune response, and lifespan in these fish, it suggests that cavefish develop these phenotypes as part of their starvation resistance and have evolved resilience phenotypes that allow them to tolerate deviations from normal vertebrate physiology. This positions cavefish as a promising model to gain mechanistic insights into disease phenotypes from an evolutionary and adaptive perspective.

Peter H. Sudmant

UC Berkeley

“Global diversity, recurrent evolution, and recent selection on amylase structural haplotypes in humans”

The adoption of agriculture triggered a rapid shift towards starch-rich diets in human populations. Amylase genes facilitate starch digestion, and increased amylase copy number has been observed in some modern human populations with high-starch intake, although evidence of recent selection is lacking. Here, using 94 long-read haplotype-resolved assemblies and short-read data from approximately 5,600 contemporary and ancient humans, we resolve the diversity and evolutionary history of structural variation at the amylase locus. We find that amylase genes have higher copy numbers in agricultural populations than in fishing, hunting and pastoral populations. We identify 28 distinct amylase structural architectures and demonstrate that nearly identical structures have arisen recurrently on different haplotype backgrounds throughout recent human history. AMY1 and AMY2A genes each underwent multiple duplication/deletion events with mutation rates up to more than 10,000-fold the single-nucleotide polymorphism mutation rate, whereas AMY2B gene duplications share a single origin.

Using a pangenome-based approach, we infer structural haplotypes across thousands of humans identifying extensively duplicated haplotypes at higher frequency in modern agricultural populations. Leveraging 533 ancient human genomes, we find that duplication-containing haplotypes (with more gene copies than the ancestral haplotype) have rapidly increased in frequency over the past 12,000 years in West Eurasians, suggestive of positive selection. Together, our study highlights the potential effects of the agricultural revolution on human genomes and the importance of structural variation in human adaptation.

Emma C Teeling

University College Dublin

"How do bats limit expected ageing and disease?"

Of all mammals, bat possess the most unique and peculiar adaptations that render them as excellent models to investigate the mechanisms of extended longevity and potentially halted senescence. Indeed, they are the longest-lived mammals relative to their body size, with the oldest bat caught being >41 years old, living approx. 8 times longer than expected.Bats defy the ‘rate-of-living’ theories that propose a positive correlation between body size and longevity as they use twice the energy as other species of considerable size but live far longer. The mechanisms that bats use to avoid the negative physiological effects of their heightened metabolism and deal with an increased production of deleterious Reactive Oxygen Species (ROS) is not known, however it is suggested that they either prevent or repair ROS damage. Bats also appear to have resistance to many viral diseases such as SARS-like and Ebola-like viruses, this suggests that their innate immunity is different to other mammals, perhaps playing a role in their unexpected longevity. Here the potential genomic basis for their rare immunity and exceptional longevity is explored across multiple bat genomes and divergent ageing and immune related markers studied in wild bat populations.

These findings provide a deeper understanding of the causal mechanisms of ageing and tolerant immunity, potentially uncovering the key molecular pathways that could be utilised to benefit society.

Donate Weghorn

Centre for Genomic Regulation

“Inference of the mutational processes and the mode of growth governing tumor evolution”

A multitude of DNA mutational processes has been identified and linked to biochemical mechanisms of DNA damage and repair. Cancer genomics relies on these so-called mutational signatures to classify tumours into subtypes, navigate treatment, determine exposure to mutagens, and characterise the origin of individual mutations. Yet, state-of-the-art methods to quantify the contributions of different mutational signatures to a tumour sample frequently fail to detect certain signatures and work well only for a relatively high number of mutations. We present a novel approach to signature decomposition using artificial neural networks that addresses these problems, allowing us to examine the origins and temporal evolution of two enigmatic and ubiquitous mutational signatures with hitherto unknown etiology. In the second part of my talk, I am going to show how we leveraged multi-region sequencing to shed light on the spatio-temporal dynamics of tumour evolution. Using two spatial metrics of evolution, we find that tumour cells grow predominantly uniformly within the tumour volume instead of at the surface. Together, our results provide new insights into the interplay between tumour biology and mutational processes as well as the constraints on the early evolution of tumours in vivo.

Jamie Blundell

University of Cambridge

“The dynamics of genetic and epigenetic alterations in the decades preceding blood cancer diagnosis”

Many human tissues are maintained by large numbers of long-lived stem cells. Positive and negative selection on somatic mutations that occur in these stem cells can drive rapid evolution in human tissues over timescales of years to decades with important implications for future cancer risk. Blood is an ideal system for gaining a quantitative understanding of these dynamics because it is easily sampled, less spatially structured relative to other tissues and genomically well characterised. Here I will describe work that exploits unique collections of serial blood samples from people destined to develop future blood cancers to reveal quantitative insights into this evolutionary process. By performing high resolution lineage tracing of genetic and epigenetic marks we are able to time the key driving events in the evolution of Acute Myeloid Leukaemia, estimate fitness effects of clones throughout the full disease trajectory and characterise competition between clones. These data shed light on the evolutionary dynamics occurring in pre-cancerous stem cells and suggest that many aspects of the observed dynamics can be understood within a surprisingly simple framework of clonal evolution with competition.

Toni Gabaldon

Barcelona Supercomputing Centre-IRB Barcelona

“Tracing the origins and recent evolution of emerging yeast pathogens”

Fungal pathogens pose a growing threat to human health. Our group has used a comparative genomics approach to study the origins and recent evolution of opportunistic pathogens in diverse yeast clades, with a focus on emerging species. The analysis of whole-genome sequences of global collections of clinical and environmental yeast isolates revealed a large worldwide genetic variation within established pathogenic species and the existence of gene flow within them. In addition it suggested multiple independent colonizations from non-human niches. Unexpectedly, a significant proportion of the analyzed isolates of rare Candida species displayed hybrid genomes, suggesting interspecies mating events. This finding challenges the classical view of stepwise pathogen adaptation and raises intriguing questions about the role of hybridization in generating novel virulence traits. Here I will present recent findings from our group on the genetic makeup and potential fitness advantages of emerging pathogens. Additionally, I will discuss the broader implications of recent genomics-enabled findings for understanding the emergence and evolution of yeast pathogens.

Aleksandra Walczak

Ecole normale supérieure-PSL

“How personalised is your immune repertoire?”

Immune repertoires provide a unique fingerprint reflecting the immune history of individuals, with potential applications in precision medicine. Can this information be used to identify a person uniquely? If it really is a personalised medical record, can it inform us about disease outcomes? I will show how statistical analysis of immune repertoires sequencing experiments can answer these questions.

Todd S. Macfarlan

Eunice Kennedy Shriver National Institute of Child Health and Human Development - NIH

“How KRAB-zinc finger proteins emerge to fight invading retroviruses in mammals”

Kruppel-associated box zinc finger proteins (KZFPs) constitute the largest family of DNA binding factors in mammals, numbering in the hundreds in most mammal genomes. They are unique amongst DNA binding transcription factors in that they have evolved rapidly throughout evolution, typically within large repetitive genomic clusters consisting of exclusively KZFP genes. Recent evidence suggests the expansion and evolution of KZFPs may be tightly linked with the endogenization of retroviruses into the germ line; as KZFPs typically bind and repress endogenous retroviruses (ERVs) of a similar age, and KZFP gene numbers correlate with the number of ERV families in different species. To gain insights into how KZFPs might respond to retroviral invasions of the germ-line, we utilized long read sequencing technologies to generate de novo assemblies of several different murine strains and sub-species, which have undergone recent infiltration of numerous ERV families.

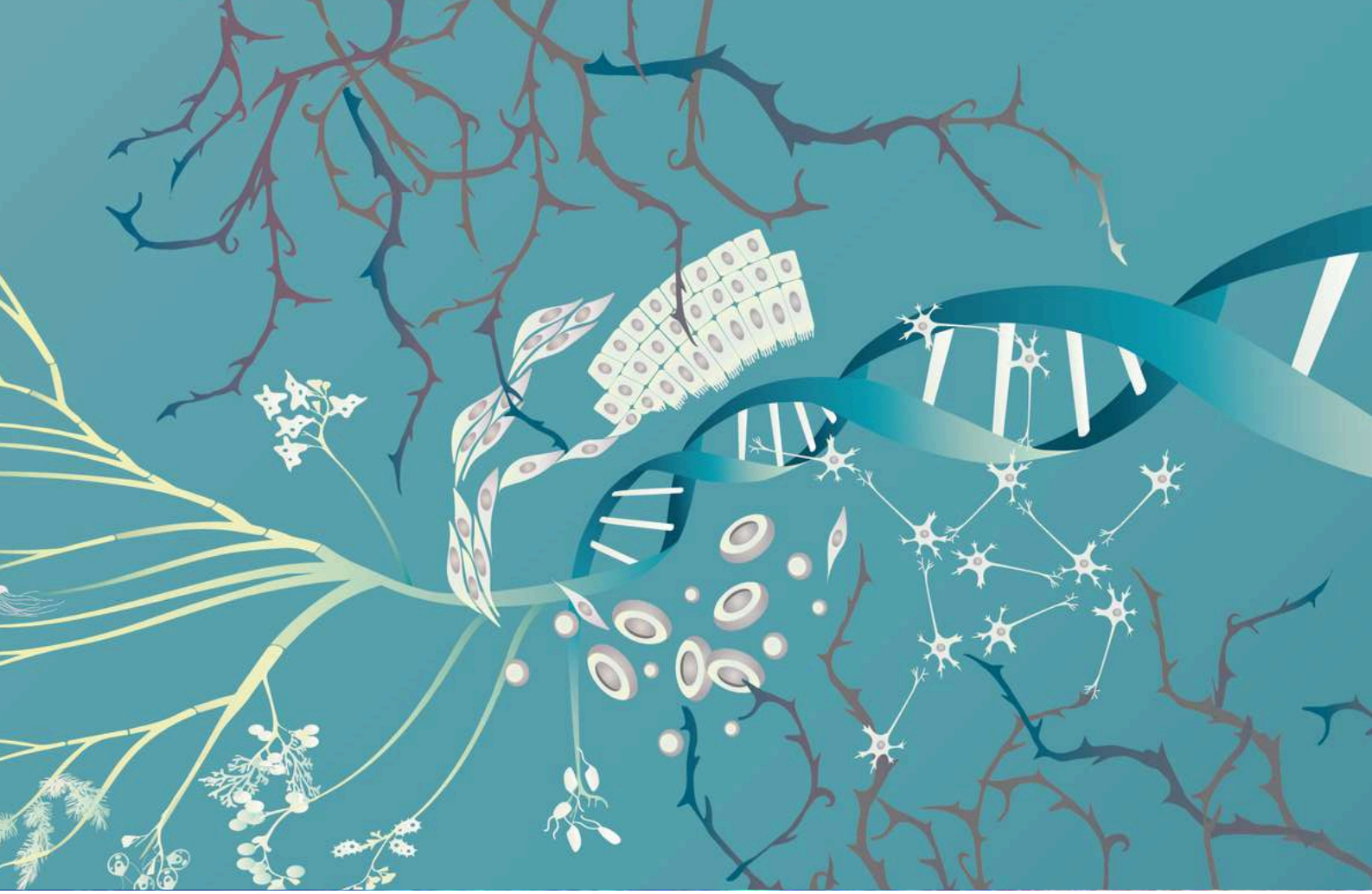
These full chromosome scale assemblies completely cover the sequence gaps in previous assemblies of the mouse genome, including previously inaccessible KZFP gene clusters. I will discuss our functional annotation of the KZFP gene clusters throughout murine species, which reveals how ERV infiltration directly into KZFP clusters facilitates KZFP heterogeneity, suggesting a direct mechanism underlying KZFP diversification in response to ERV activity.

Mireia Valles Colomer

University Pompeu Fabra

“Within host evolution and transmission dynamics of the human microbiome”

The human microbiome (the collection of microorganisms in and on our bodies) is an integral component of the human body and a co-determinant of multiple diseases. Being modifiable with non-invasive interventions, it holds potential for the development of therapeutic approaches. However, the persistence and evolution rates of microbiome components together with their transmission dynamics have been until recently largely unexplored. The first members of the microbiome are acquired from our mothers at birth and soon thereafter, and in the largest study of microbiome interhost transmission to date we found microbiome seeding to be complemented by proximate individual’ microbiomes. Host genetics only displayed a moderate effect on interhost transmission dynamics, and within host evolution rates were most strongly associated with environmental factors. Finally, we observed distinct alterations in transmission patterns in microbiome-associated diseases.



Contributed talks

The abstracts can be found in the overview with the selected posters

Poster 1 **Ariadna Bada i Navarro**

“Genomic Insights into the Evolutionary Nature of Autism Spectrum Disorder”

Poster 10 **Genevieve Housman**

“Interrogating osteoarthritis-related regulatory dynamics in primates using skeletal cell culture systems”

Poster 11 **Konstantinos Karakostis**

“Evolution of the p53/MDM2 interface leads to gain of functions in development and tumour suppression”

Poster 15 **Yosuke Nagahata**

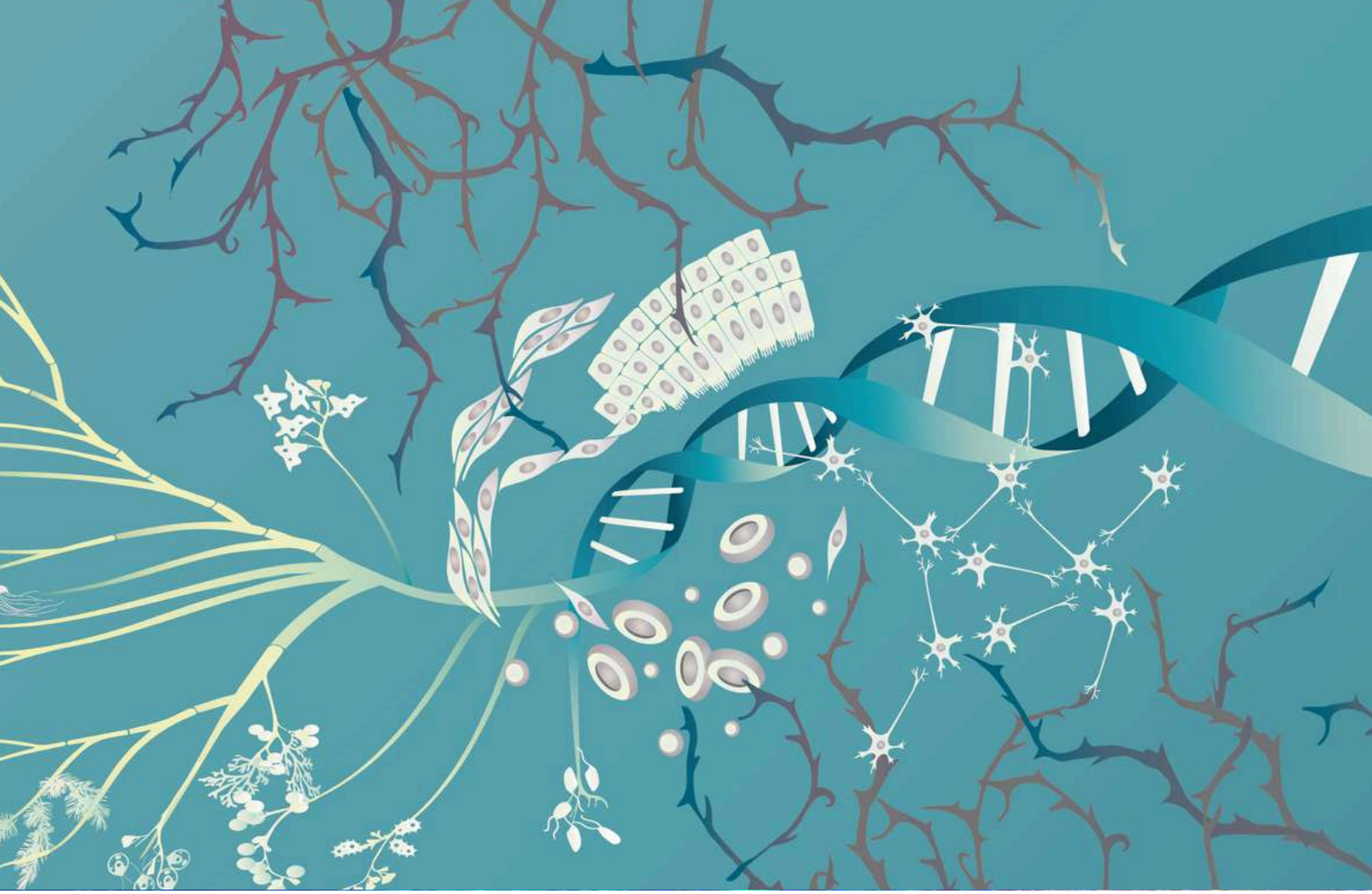
“Evolutionary reversion in tumorigenesis”

Poster 19 **Irepan Salvador**

“Adaptation in Human immune cells residing in tissues at the frontline of infections”

Poster 20 **Miquel Àngel Schikora Tamarit**

“Machine learning classifiers reveal the evolutionary drivers of virulence and drug resistance in the fungal pathogen Candida parapsilosis”



SELECTED POSTERS

Poster 1

Ariadna Bada i Navarro

University of Barcelona

"Genomic Insights into the Evolutionary Nature of Autism Spectrum Disorder"

Background:
Given the impairing nature and reproductive drawbacks of Autism Spectrum Disorder (ASD), its high prevalence in the population poses an evolutionary paradox. In principle, natural selection removes alleles that decrease fitness of the carriers – i.e. risk alleles with large effects on predisposition to mental illnesses. However, a substantial fraction of ASD genetic risk is attributable to a polygenic predisposition driven by common variation of small to moderate effect size. We aimed to investigate the evolutionary nature of ASD from a genomic perspective applying different state-of-the-art evolutionary genomic tools, using the most recent ASD-GWAS meta-analysis including 38,717 cases and 232,735 controls (unpublished).

Methods:
We used MAGMA to test whether ASD-associated variants are enriched in LoF-intolerant (Leeuw et al. 2015) and HARs (Girskis et al. 2021) genes. We tested whether ASD-risk alleles (stratified by MAF bins) tend to be enriched for the derived or ancestral allele (present in chimpanzee (Pan Troglodytes)).

Poster 1

Moreover, we investigated the load of ASD□risk-alleles in ancient genomes from the Allen Ancient DNA Resource (AADR) (Mallick et al. 2024).

Results:
MAGMA analyses suggest that ASD-associated variants are enriched in LoF-intolerant genes (ExAC pLI≥0.9) (p=6.64e-08). This enrichment persists in LoF-intolerant brain□expressed (FPKM≥1, p=2.99e-07) and highly-brain-expressed (FPKM≥5, p=5.44e-07) genes. ASD-associated variants are also enriched in HARs genes (p=0.0005). Our results indicate that the ancestral alleles generally reduce susceptibility to autism compared to the derived state. Trend analysis of ancient DNA samples revealed a statistically significant increase in the proportion of ASD-risk alleles towards the present (τ=-0.099, p=0.034)

Discussion:
These results, together with new genomic approaches that we are conducting using archaic genomes could hint at the evolutionary nature of ASD, providing a new dimension for understanding this disorder. Further analyses to study the pleiotropy of ASD-associated variants with other phenotypes will help us to unravel the selective forces shaping ASD

Poster 2

Pierre Bercier

CRG

"The “Gelling” Cure: How Arsenic Trioxide Targets PML Bodies to Cure Leukemia"

Acute promyelocytic leukemia (APL) is driven by the PML-RARA oncoprotein, which disrupts the assembly of PML nuclear bodies (PML NBs)—key biomolecular condensates involved in essential cellular functions. Arsenic trioxide (ATO) has emerged as a highly effective treatment, curing 70% of patients as a monotherapy and up to 98% in combination therapies. Despite this success, the molecular mechanisms underlying ATO’s curative effects were not fully understood. Through an international collaboration, we uncovered the molecular basis of ATO’s therapeutic action. ATO binds to the PML protein, inducing a liquid-to-gel transition in PML-RARA microspeckles. This transition triggers the degradation of the PML-RARA oncoprotein, leading to the elimination of leukemic cells. Interestingly, ATO also induces the gelling of normal PML NBs. Our research focused on the B-box-2 (B2) domain of the PML protein, mutated in ATO-resistant APL patients. By elucidating the crystal structure of B2, we identified a key α-helix that mediates B2 trimerization through hydrophobic interactions, controlling PML NB assembly and dynamics. B2 trimerization positions a critical cysteine triad in its center, forming an ideal docking site for arsenic.

Poster 2

Arsenic binding to this cysteine triad induces a structural shift of PML NBs from a liquid-like to a gel-like state, which is crucial for initiating PML-RARA degradation and ultimately, APL cure. This study presents a pioneering example of targeted therapy against nuclear condensates, demonstrating that altering their material properties can directly result in disease eradication. The insights gained offer promising avenues for treating cancers that involve PML dysregulation.

Poster 3

Federico Billeci

CRG

"Evolutionary-informed probabilistic models for genetic disease discovery"

Polygenic disorders are a broad class of diseases caused by multiple genetic mutations, often involving noncoding and regulatory variants. The most common approaches to predicting phenotype from genotype are scores computed with simple regression models but their potential as diagnostic tools is still heavily debated. Indeed, these models provide very limited insight into disease aetiology. This is largely because current approaches consider only common variants and are unable to identify the relationship between variants and the pathways they affect. We propose a probabilistic and generative modelling framework that predicts the joint impact of both common and rare variants and reveals new disease sub-classes. Additionally, our approach leverages evolutionary constraints to predict variants' pathogenicity. Indeed, recent evidence has shown that sequence variation across organisms offers a very accurate guide for disease-causing variants in severe developmental disorders, making evolutionary information an essential tool for studying genetic diseases. Still, the potentiality of evolutionary constraints for predicting polygenic phenotypes and explaining their strong heterogeneity remains mostly unexplored. We apply this modelling approach to type 2 diabetes, a disease with large genetic heterogeneity, for which clear diagnostic criteria are still lacking

Poster 4

María Elena Campoy García

Hospital del Mar Research Institute

"Towards a complete characterisation of human polymorphic inversions and their functional effects"

Structural variants (SVs) contribute substantially to genetic and phenotypic diversity, but their characterization is far from complete. Inversions are particularly interesting because they affect recombination and could have negative consequences on fertility. However, they are often missed due to their balanced nature, the repetitive sequences at their breakpoints and the fact that many are recurrent. Here, thanks to an in-depth analysis of >350 predictions from different studies, manual annotation, and accurate validation and genotyping in diverse populations, we have generated the largest and most reliable dataset of human polymorphic inversions to date. This unique resource totals 134 inversions and 61 inverted duplications, making it finally possible to determine their real functional and evolutionary impact. In particular, by rigorous imputation in available functional data, we have shown that ~25% of the studied inversions act as lead expression QTLs (eQTL) for different genes in multiple tissues. Moreover, several inversions are associated with epigenetic changes in chromatin accessibility, DNA methylation or histone marks. Finally, ~20% of inversions are in high linkage disequilibrium (LD) with GWAS signals, including an inversion with clear frequency differences across continents that is associated to body shape and height.

Poster 4

Remarkably, when compared to SNPs, inversions tend to be enriched in functional effects, especially the largest ones that have already been proposed to act as supergenes, which could compensate their potential fertility costs. In fact, we detected potential signatures of natural selection acting on these variants, as evidenced by elevated FST values among populations and higher-than-expected frequencies of NAHR inversions when compared to neutral SNPs. Therefore, these findings highlight the important role that inversions can play in many organisms and reveal previously missing variants responsible for human phenotype variability

Poster 5

Renata Cunha

UPF

"Impact of IsletMICs deregulation in islet development"

Islet of Langerhans are micro-organs that comprise around 1-2% of all pancreatic tissue and are responsible for regulating glucose homeostasis through the secretion of insulin and glucagon. Islet dysfunction plays a central role in diabetes, and the spatial architecture of islet cell types is crucial for proper hormone secretion. Despite significant knowledge regarding islet cell differentiation and mature secretory function, little is known about the regulatory programs controlling islet morphogenesis and establishment of cytoarchitecture. Recently, we identified a novel program of microexons in islets (IsletMICs), impacting genes related to insulin secretion and type 2 diabetes. Microexons (3 to 27 nucleotides) are very small exons regulated by RNA binding proteins SRRM3 and SRRM4, originally described as a conserved program in neurons playing crucial roles in neural development and function. Islet splicing, and particularly IsletMICs, is enriched in genes involved in cell adhesion and cell migration, suggesting that alternative splicing plays important roles in islet development. To study islet organogenesis, we developed constitutive and conditional SRRM3-knockout mice lines. Immunostaining of pancreas from SRRM3-knockout mice shows altered islet architecture and cell type composition, with loss of the stereotypical mantle-core design.

Poster 5

In addition, these mice show islet functional defects, causing impaired insulin release and glycaemic alterations. We hypothesise that developmentally-regulated islet splicing, specially IsletMICs, regulates islet development by affecting islet morphogenesis and the establishment of a mature architecture. Deregulation of microexon inclusion may influence endocrine cell fate, differentiation and cell-cell communication, thereby disrupting islet organisation and contributing to impaired hormone secretion and glucose homeostasis.

Poster 6

Thomas Dupic

Harvard FAS

"Parallel Evolution of the Antibody Response during SARS-CoV-2 Vaccination"

We longitudinally sequenced, over multiple months, the antibody repertoires of 10 individuals following SARS-CoV-2 infection, then vaccination, generating a dataset of 10^8 lineages. We then selected a subset of 10’000 potentially reactive antibodies and measured their affinity against the spike protein of the virus. To identify relevant antibodies, we modeled the statistical properties of each repertoire, revealing striking similarities across individuals. We then selected lineages that appeared to be outliers —either because they were shared across multiple individuals more frequently than expected, or because they exhibited important changes in frequency or phylogenetic structure over time. Our analysis revealed patterns of convergent evolution, with similar naive antibodies displaying parallel evolutionary trajectories across individuals. Additionally we developed and tested methods to identify reactive antibodies within repertoire datasets. Our findings offer insights into the shared evolutionary dynamics of the antibody responses and contribute to improving our strategies for analyzing immune repertoires.

Poster 7

Albert Escobedo

CRG

"Genetics, energetics and allostery during a billion years of hydrophobic protein core evolution"

Protein folding is driven by the burial of hydrophobic amino acids in a tightly-packed core that excludes water. The genetics, biophysics and evolution of hydrophobic cores are not well understood, in part because of a lack of systematic experimental data on sequence combinations that do - and do not - constitute stable and functional cores. Here we randomized protein hydrophobic cores and evaluated their stability and function at scale. The data show that vast numbers of amino acid combinations can constitute stable protein cores but that these alternative cores frequently disrupt protein function because of allosteric effects. These strong allosteric effects are not due to complicated, highly epistatic fitness landscapes but rather, to the pervasive nature of allostery, with many individually small energy changes combining to disrupt function. Indeed both protein stability and ligand binding can be accurately predicted over very large evolutionary distances using additive energy models with a small contribution from pairwise energetic couplings. As a result, energy models trained on one protein can accurately predict core stability across hundreds of millions of years of protein evolution, with only rare energetic couplings that we experimentally identify limiting the transplantation of cores between highly diverged proteins. Our results reveal the simple energetic architecture of protein hydrophobic cores and suggest that allostery is a major constraint on sequence evolution.

Poster 8

Manuela Giraud

CRG

"Long Insertions and Deletions (InDels) contribute both to robustness and evolvability in Genotype-Phenotype maps"

Changes in phenotypic traits in evolution must originate from random mutations in the genotype. Thus, it is important to understand the mapping from genotype to phenotype (GP map), where phenotypes are higher order structures that can be considered at many levels (amino-acids, proteins, reaction networks ...). Previous studies have looked in depth at the properties of GP maps, including phenotype robustness, frequency and evolvability, and how they relate to concepts like epistasis. These properties play a part in determining which new phenotypes can arise and be selected for. However, these analyses have focused almost exclusively on substitutions. Other types of mutations like insertions and deletions (InDels) are less well-understood, even though they are ubiquitous and may play fundamentally different roles within a GP map, for example by providing access to qualitatively different structural variation. Here, we use an established computational GP map model based on RNA secondary structure folding to study InDels of lengths l = 1 to l = 5. We find high mutational robustness even for l = 5 InDels compared to null models, such as comparing l = 5 indels to five separate l = 1 indels. This high neutrality may allow some InDels to fix without fitness penalty, and act as a stepping stone to phenotypic variation.

Poster 8

On the other hand, InDels that do change the phenotype, enable phenotypic changes not possible through substitutions. Together, these results imply that InDels, even if rare, play an important part in neutral and non-neutral evolution on GP maps.

Poster 9

Noemí Hostalet

FIDMAG Germanes Hospitalàries Research Foundation

"Human-Specific Evolutionary Markers and Sulcal Pits patterns: Implications for Schizophrenia Neurodevelopmental trajectories"

Background:
Schizophrenia may represent a trade-off in the evolution of human-specific neurodevelopmental mechanisms. Human Accelerated Regions (HARs) are evolutionary markers functioning as neurodevelopmental transcription enhancers, associated with brain configuration, neural information processing, and schizophrenia risk. We aimed to investigate whether the HARs polygenic load influences neuroanatomical characteristics in schizophrenia, focusing on sulcal pits – deepest points of the cortex - as potential markers for understanding human neurodevelopmental trajectories and related disorders.

Methods:
The sample comprised 189 healthy controls (HC) and 237 age-/sex-matched schizophrenia patients with T1-weighted MRIs and genome-wide genotyping. Sulcal pits were identified from white matter surfaces generated by FreeSurfer. Graphs were created for each hemisphere and lobe, treating sulcal pits as nodes and their connections as edges, and were compared based on area, pit depth, pit position, graph topology, and combined features.

Poster 9

Similarity scores were calculated by comparing each participant to all HC. Using the latest schizophrenia GWAS, we computed polygenic risk scores (HARs-PRSsz). We then: i) compared mean similarity scores between HC and patients, and ii) analyzed the effect of HARs-PRSsz on the similarity of patients' sulcal pits graphs.

Results:
Patients with schizophrenia displayed a differential sulcal pits pattern (combined feature) in the frontal lobe, showing reduced similarities as compared to HC ($t = 3.01$; $p = 0.002$; $d = - 0.23$). Among patients in the highest HARs-PRSSZ quartile, greater deviations from the HC sulcal pits pattern of reference were associated with an increased polygenic burden ($R^2 = -0.46$; $p = 0.02$)

Conclusions:
The results support the early neurodevelopmental component to the disorder-specific sulcal pattern identified in the left frontal lobe, correlated with HARs-PRSsz. This is aligned with previous reports of this region likely being influenced by human-specific evolutionary markers.
Acknowledgements: NARSAD Award (26206-MFV); ISCIII project-PI20/01002, grant□FI21/00093-NH, grant-CP20/00072-MFV and mobility-MV23/00059-NH; co-funded by ERDF/ESF); AGAUR-2021SGR01475

Poster 10

Genevieve Housman

Max Planck Institute for Evolutionary Anthropology

"Interrogating osteoarthritis-related regulatory dynamics in primates using skeletal cell culture systems"

Purpose:
Primates display varied susceptibilities to skeletal pathologies, like the chronic joint disease osteoarthritis (OA) which is highly prevalent in humans and some cercopithecoids and minimally prevalent in nonhuman apes. These differences may be due to genetic factors, environmental factors, or a combination of both. Such factors may mediate their phenotypic effects though changes in gene regulation, but isolating which mechanisms contribute to OA variation across species is a challenge. This project uses comparative primate skeletal cell culture systems to begin dissecting OA-related regulatory dynamics across different genetic backgrounds and in response to trait-relevant environmental exposures.

Methods:
A panel of 6 human and 6 chimpanzee induced pluripotent stem cells were differentiated into chondrogenic cells and exposed to either control conditions or mechanical stress and inflammatory cytokine treatments that induce a matrix-degrading, OA-like phenotype. RNA-sequencing data were collected across individuals and conditions in order to measure gene expression variation, and replicate experiments were performed for 3 humans and 3 chimpanzees to assess technical variation.

Poster 10

Results:
Gene expression patterns in technical replicates were strongly correlated ($r>0.91$), validating the reproducibility of cell differentiations and OA-like treatment conditions.
Most genes show conserved expression levels between humans and chimpanzees, with only a portion of genes differentially expressed between species. Treatments show smaller effects on gene expression, but when exposed to the same OA-related environmental factors, humans and chimpanzees display a mixture of shared and unique gene expression responses.

Conclusions:
The mixture of conserved and divergent OA-related regulatory dynamics in humans and chimpanzees – phylogenetically-close taxa that display different natural prevalence rates of OA – provide some insight into the factors contributing to evolutionary differences in OA susceptibility. Continued work using these skeletal cell culture systems in larger panels of individuals, species, and relevant conditions will help to further clarify and disentangle these mechanisms.

Poster 11

Konstantinos Karakostis

IMIM

"Evolution of the p53/MDM2 interface leads to gain of functions in development and tumour suppression"

The p53 transcription factor is a key cell cycle regulator and a powerful tumour suppressor. Throughout evolution, the ancestral p53/63/73 gene having developmental roles in cell differentiation evolved into three genes with district functions in development, cancer, ageing and stress responses. P53 is mainly regulated by the MDM2 E3 ubiquitin ligase and ATM via the p53boxI motif. Depending on DNA damage conditions, mammalian MDM2 alternatively binds the p53 protein (degradation) or p53boxI mRNA (activation). However, the evolution of these interactions remains understudied. Here we aim to understand p53/MDM2 structural interfaces and signalling mechanisms to establish how p53 evolved in the context of co-occuring exogenic stresses to gain functions in development and cancer. The MDM2-p53 protein–mRNA and protein–protein interactions were studied in humans, invertebrates (Ciona intestinalis, Ci) and in elephants, which are unique in having >20 p53 retrogene copies. Methods included in silico modelling, in vitro and cellular assays to solve the p53 mRNA structures and the signalling mechanisms. We found that a temperature-sensitive p53 mRNA structure in invertebrates controls its interaction with MDM2. This interaction evolved to become regulated by the ATM kinase in mammalian DNA damage responses, leading to p53 activation.

Poster 11

Also, a non-conserved flanking region of Ci-p53boxI prevents the p53-MDM2 protein–protein interaction, indicating that negative regulation of p53 by MDM2 evolved in vertebrates. Therefore, from invertebrates to mammals, p53 mRNA and protein structures evolved independently to interact with MDM2. Finally, partial p53 proteins encoded by p53 retrogenes uniquely found in the elephant lineage, distinctively bind MDM2 in a temperature-dependent manner, thus influencing the activation of canonical p53 in response to genotoxic or heat stress. These results address the evolutionary interface of environmental conditions impacting on cell regulation, giving insights to targeted therapeutic strategies.

Poster 12

Junho Lee

Korea Institute of Science and Technology

"Role of cancer associated fibroblasts in the tumor microenvironment: mathematical modeling"

The heterogeneity of cancer-associated fibroblasts (CAFs) within the tumor microenvironment (TME) plays a pivotal role in the progression and treatment of cancer. Understanding the distinct behaviors and effects of various CAF phenotypes is crucial for the development of more effective cancer treatment strategies. This study is driven by the purpose of elucidating the heterogeneity of CAFs within the TME and evaluating how different CAF phenotypes influence tumor progression and immune response dynamics. However, the exact role and mechanism of CAFs within the TME remains to be elucidated. This study highlights the need to dissect the complex roles of different CAF phenotypes that affect cancer progression and immune regulation and proposes mathematical models to explore these interactions. Utilizing method that combine an agent-based model with differential equations, we simulate the complex interactions between CAFs and T cells and analyze spatial effect. The agent-based modeling allows for the simulation of individual cellular behaviors within their microenvironments. This method provides nuanced insights into how cells interact with and influence their surroundings, which is critical for understanding the complexity of the TME.

Poster 12

The results of our simulations indicate that the anti-immune CAF phenotype contributes to the inhibition of T cell activation, thereby enhancing tumor survival, while the prophylactic immune phenotype may support T cell activity and thus disrupt tumor growth. The conclusion of our study suggests that strategic manipulation of the CAF population may significantly alter the immune environment of TME, suggesting new avenues for treatment of target cancer. Adjusting CAF phenotypes can improve the effectiveness of immunotherapy strategies and lead to more personalized approaches in cancer treatment.

This study not only advances our understanding of the TME's complexity but also opens new avenues for the development of personalized cancer therapies that are finely tuned to the specific characteristics of the tumor environment.

Poster 13

Ettore Luzi

UNIFI

"Perturbation of microrna regulatory networks design different waddington's epigenetic landscapes for neuroendocrine tumors"

Cancer is a heterogeneous progressive disease caused by perturbations of the underlying Gene Regulatory Networks (GRNs), particularly by sub networks that appear frequently and are functionally important (i.e., GRN motifs). that can be described by dynamic models. TFs and microRNAs act in concert in networks to regulate target genes in a coordinated manner. TFs and microRNAs are in turn regulated, in part, at transcriptional and post-transcriptional levels. In line, regulatory nodes may comprise TFs and miRNAs that form sub-networks including fundamental, evolutionary conserved regulatory motifs such as feedback or feed forward loops (FBL, FFL). A classical GRN motif, the toggle switch, constitutes a molecular mechanism that determines cell-fate decisions, and provides stability to transcriptional programs of binary cell-fate choices. Over-expression of each transcription factor (TF)-microRNA-mRNAs corresponds to one of the two mutually exclusive cell fates normal and disease states of the cell) whereas a ‘‘balanced’’ expression of both TFs-microRNAs -mRNAs maintains the normal state.

Poster 13

Here, through a global human miRNA next generation sequencing (NGS) screening, we analysed the expression differences between healthy pancreas, pancreatic neuroendocrine tumor, and duodenal gastrinoma : The analysis showed miRNAs “hubs” organized in FFLs with chromatin-remodeling genes (MEN1, ATRX, DAXX) involved in familial and sporadic cancer, as well as with classical oncogenes (RB, TP53) normally mutated in neuroendocrine carcinomas (NEC) suggesting that the transition from normal neuroendocrine cells to well-differentiated neuroendocrine cancer or to a poorlydifferentiated neuroendocrine carcinoma could be controlled by this epigenetic miRNAbistable molecular regulatory circuit. Following Conrad Waddington’s epigenetic landscape of cell development, it is possible to use a Hopfield network formalism to construct an attractor landscape model of disease progression.

Poster 14

Selena Aranda

Universitat de Barcelona

"Investigating the link between social traits and Human Accelerated Regions in the genome"

Introduction:
Sociability is a poorly understood trait influenced by environmental and genetic factors. Some patients with schizophrenia present social impairment, which significantly impacts their quality of life and daily functioning. Understanding the genetic basis of sociability could lead to personalized therapies targeting social dysfunction, improving outcomes and social integration. Human Accelerated Regions (HARs) are genomic regulatory elements that have undergone rapid evolutionary changes since human-chimpanzee divergence, suggesting they may play a crucial role in what makes us uniquely human. Thus, we aimed to explore the link between HARs and sociability using different genomic approaches.

Methods:
Sociability-associated loci were retrieved from the largest GWAS on sociability (N=342,461 participants, UKBiobank; Bralten et al. 2021), and the list of HARs from Girskis et al. 2021. We investigated whether sociability-associated SNPs lie near HARs by comparing the observed mean distance between sociability-associated SNPs and their nearest HAR against a null distribution of randomly selected SNPs.

Poster 14

MAGMA was used to test if sociability-associated SNPs are enriched in genes near HARs (de Leeuw et al. 2015). Additionally, we calculated the global Polygenic Score (Global PGS) and HAR-specific PGS for sociability using PLINKv1.9 (Purcell et al. 2007) and studied their association with the PANSS Autism Severity Score (PAUSS; Kästner et al. 2015) in N=553 schizophrenia patients (PsyCourse; Budde et al. 2018).

Results:

Sociability-associated SNPs were found to be near HARs ($p<10^{-5}$) and enriched in genes located near HARs ($p=3.7x10^{-4}$). However, neither Global PGS nor HAR-specific PGS showed an association with PAUSS in schizophrenia patients.

Discussion:

Our results highlight the involvement of HARs in social traits. However, preliminary analyses failed to link HARs' polygenic burden with social impairment in schizophrenia. We are currently studying HARs' polygenic load on social traits in the general population and other psychiatric disorders aimed at understanding the link between HARs and social traits

Poster 15

Yosuke Nagahata

IBE (CSIC-UPF)

"Evolutionary reversion in tumorigenesis"

Introduction:
Cells forming malignant tumors are distinguished from those forming normal tissues based on several features: accelerated/dysregulated cell division, disruption of physiologic apoptosis, maturation/differentiation arrest, loss of polarity, and invasive potential. Among them, accelerated cell division and differentiation arrest make tumor cells similar to stem / progenitor cells, and this is why tumorigenesis is often regarded as developmental reversion. Here, in addition to developmental reversion, we propose another insight into tumorigenesis from a phylogeny viewpoint. Because tumor cells also share some features with unicellular organisms, we propose that tumorigenesis can be regarded as “evolutionary reversion”.

Methods:
Recent advances in sequencing technologies and the ability to identify homologous genes have made it possible to perform comprehensive cross-species transcriptome comparisons. By such comparisons, we recently traced the evolutionary history of blood cells back to the unicellular ancestor of animals (Nagahata et al, Blood, 2022, Nagahata et al, Front in Oncol, 2023). In this study, by expanding the methodology, we compared transcriptome data of tumor cells, normal cells, and various unicellular organisms.

Poster 15

Results:
We found that malignant tumor cells transcriptionally resembled unicellular organisms. In addition, it was also suggested that similarities between tumor cells and unicellular organisms were more clear in eukaryotic unicellular organisms rather than prokaryotes, and also increased along with tumor progression: from normal cells to premalignant cells, primary tumor cells, and metastatic tumor cells. Furthermore, we found that similarity to unicellular organisms improved risk-classification in leukemia clinical data; patients with unicellular eukaryote-like leukemia cells showed worse survival rate, while those with prokaryote-like leukemia cells showed better survival rate.

Conclusion:
Analyzing tumorigenesis from the viewpoint of phylogeny should improve the therapeutic strategies, and can reveal new aspects of tumorigenesis to overcome malignant tumors.

Poster 16

Colette Nickodem

University of Wisconsin - Madison

"Linking mobile genetic elements to the transmission of antimicrobial resistant cattle pathogens and their impact on farmer micro"

Objective:
The goal of this research is to determine the contribution of on-farm animal exposure and biocide usage to changes in the human microbiome. We explored levels of AMR, and risks of harboring and transmitting drug resistant pathogens, that contribute to gut and respiratory dysbiosis and other health complications.

Methods:
Participants from an ongoing population-based cohort study answered a comprehensive health survey and provided stool samples for microbiome analysis. Amplicon sequencing was conducted on stool samples from farmworker (n=44), healthcare worker (n=44), and control (n=44) populations to explore microbiome changes based on workplace exposures. Farmworker stool and nasal swabs were also submitted for shotgun sequencing for an in-depth exploration of ARGs and MGEs. In a separate study, fecal samples from cattle and farm workers (n=50), at high and low antibiotic use farms, underwent shotgun sequencing. Sequencing files from both studies were aligned using AMR++ v3.0 to obtain a baseline representation of taxonomy and ARG presence.

Poster 16

Resistome risk scores were obtained using MetaCompare 2.0 and high-quality metagenome assembled genomes (MAGs) were created to conduct phylogenetic analyses for exploration of AMR pathogen evolution.

Results:
The farmworker population from the cohort study had a higher burden of chronic diseases, in comparison to healthcare workers and non-farm or healthcare workers. Rates of myocardial infarction, osteoarthritis, rheumatoid arthritis, and stroke were drastically higher in the farm population. Due to the connection to respiratory and gut microbiome, the farm population contributed to 40% (6/15) of asthma cases and 33% (2/6) of diabetes cases. A higher proportion of zoonotic pathogens were identified within the farmworker population.

Conclusions:
Parsing sequences for ARGs and MGEs in close proximity within pathogens will determine the risk of cross species transmission. Integrating the shotgun sequencing results between these two studies will help us disentangle the complex interactions between farmworkers and pathogens within their workplace environment.

Poster 17

Sofia Papanikolaou

University of Crete

"Transcriptional dynamics and alternative splicing events in Systemic Lupus Erythematosus"

Systemic Lupus Erythematosus (SLE) is a complex autoimmune disorder characterized by production of autoantibodies against nuclear and cytoplasmic antigens and multisystem inflammation. Previous studies revealed that genes implicated in disease pathogenesis are subject to alternative splicing (AS)1,2. In addition to increasing the complexity of the transcriptional output, alternative RNA splicing can lead to the reduction of mRNA translation or the production of non-functional or malfunctional proteins, thus representing a vital component of gene regulation process3. The aim of our study was to identify and characterize alternative splicing events in individuals with Systemic Lupus Erythematosus (SLE) compared to healthy counterparts. This analysis encompassed a range of specific cell types as well as whole blood samples from patients exhibiting variable levels of disease activity. Through the implementation of a computational pipeline on publicly available and our own RNA-sequencing data, we performed differential splicing analysis and uncovered significant alterations in transcription dynamics that affected a substantial number of genes. Alternative splicing impacted a distinct set of genes as compared to those identified as differentially expressed.

Poster 17

Differential splicing analysis between SLE versus healthy individuals revealed substantial cell type-specificity of splicing events. We observed a prevalence of intron retention events, with the majority leading to the introduction of premature stop codons, implying potential gene repression. Notably, intron retentions were detected in transcripts of healthy samples and this phenomenon was diminished in the context of SLE. We examined intrinsic properties of the retained introns, including their GC content and length. Functional assessment of genes influenced by alternative splicing pointed towards specific roles in metabolism and histone acetylation as areas of potential significance. Overall, our findings emphasize the critical role of incorporating alternative splicing analyses in the molecular characterization of complex diseases like SLE.

Poster 18

Charles Pugh

CRG

"Investigating the dynamics of protein constraint across the tree of life with deep Bayesian hierarchical models"

Generative probabilistic models of protein sequences are transforming protein design, pathogen forecasting and clinical variant annotation. In the latter case, provided certain criteria are met, these models can now be considered strong evidence for missense variant pathogenicity and are an invaluable diagnostic tool. Crucially to their success, these expressive models are able to capture constraints on protein sequences by learning from variation across the tree of life. However, current variant effect prediction approaches model the distribution of broad protein families or, in the case of protein language models, the entire protein universe in order to make predictions in a specific species. Modelling all sequences as from a single distribution in this way overlooks the changes in constraint on a protein across the tree of life. To address this issue we have developed a novel modelling strategy to capture the distribution of sequences conditioned on the position in the phylogeny. We achieve this with a deep Bayesian hierarchical model based on the phylogenetic structure of large multiple sequence alignments. Our model learns from diverse sequences across the phylogeny while also capturing constraints that have evolved more recently. This has enabled us to obtain state-of-the-art scores on the ProteinGym benchmark.

Poster 18

Our approach has allowed us to trace the evolutionary history of constraints in a selection of disease relevant human proteins. With this we have identified differing evolutionary signatures in gain-of-function and loss-offunction variants, giving promise that our approach can help classify disease aetiology.

Poster 19

Irepan Salvador-Martínez

CNAG

"Adaptation in Human immune cells residing in tissues at the frontline of infections"

Human immune cells are under constant evolutionary pressure, primarily through their role as first line of defence against pathogens. Most studies on immune adaptation are, however, based on protein-coding genes without considering their cellular context. Here, we combine a population genomics approach and data from the Human Cell Atlas to infer the gene adaptation rate of the human immune landscape at cellular resolution. Our analysis revealed abundant cell types, like progenitor cells during development and adult cells in barrier tissues, to harbour significantly increased adaptation rates. Leveraging spatial transcriptomics data, we confirmed the adaptation of tissue-resident T and NK cells in the adult lung located in compartments directly facing external challenges, such as respiratory pathogens. Finally, by analyzing human iPSC-derived macrophages responding to various challenges, we found adaptation in early immune responses. Together, our study suggests host benefits to control pathogen spread at early stages of infection, providing a retrospect of forces that shaped the complexity, architecture, and function of the human body.

Poster 20

Miquel Àngel Schikora Tamarit

Barcelona Supercomputing Centre - IRB Barcelona

"Machine learning classifiers reveal the evolutionary drivers of virulence and drug resistance in the fungal pathogen Candida parapsilosis"

Fungal pathogens pose a serious health threat, causing ~4 million yearly deaths,requiring improved therapeutic and diagnostic tools. Candida species are major contributors to hospital-acquired infections, particularly among immunocompromised patients. Given the high dynamism of Candida genomes, a promising strategy to improve current therapies and diagnostics is to elucidate the evolutionary mechanisms generating virulence and antifungal drug resistance. These remain obscure because previous studies had small sample sizes, lacked rigorous statistical analyses, focused only on expected genes and/or ignored the role of genetic interactions.

Here, we investigated such evolutionary processes in Candida parapsilosis, an emergent pathogen frequently causing nosocomial outbreaks. We sequenced the genomes and measured several phenotypes (e.g. drug susceptibility, invasiveness) for hundreds of clinical isolates from six Spanish cities. This large dataset enabled unprecedented statistical power to study drug resistance and virulence. For this, we first performed a convergence genome-wide association study to pinpoint variants (including structural variants) underlying these traits.

Poster 20

Additionally, we used explainable machine learning classifiers (e.g. based on random forests) to predict each phenotype from genetic variants.

Our analyses revealed hundreds of potential genetic drivers of virulence and resistance, confirming expected evolutionary mechanisms, and suggesting novel ones. We were able to build highly accurate classifiers for some phenotypes. For instance, we can predict changes in fluconazole resistance, the mainly used drug for C. parapsilosis, with high confidence. Our results provide insights on the emergence of clinically-relevant traits in a major fungal pathogen, and set the foundations for sequence-based characterization of antifungal susceptibility in the clinics.

Authors: Miquel Àngel Schikora-Tamarit, Óscar Zaragoza, Toni Gabaldón

Poster 21

Xavier Soler

CRG

"Beyond Driver Mutations: Evolutionary Clustering of Cancer Patients for Precision Medicine"

Cancer involves a number of different changes that lead to abnormal and uncontrolled cell growth. Analogous to Darwinian evolution in the origin of species, cancer development is based on two constitutive processes: the continuous acquisition of heritable genetic variation in individual cells by more or less random mutation, and natural selection acting on the resulting phenotypic diversity. This results in a high mutational, biochemical, and histological tumor heterogeneity that makes driver mutation identification very challenging. Historically, patients have been clusterized according to their mutated driver genes or driver mutations. In many occasions, this clustering does not go hand in hand with their treatment response. The arrival of second-generation sequencing technologies has allowed us to produce a huge amount of data, including enormous amounts of whole genomes from cancer patients. And this is following an ongoing increase since the sequencing costs are continuously decreasing. Together with this trend, we are seeing the emergence of personalized medicine, which aims to maximize the value of data from individual patients. Our approach exploits the power of this data to cluster patients by tumor mechanisms.

Poster 21

We use whole genome information combined with the functional impact of the mutations learned from genetic variation seen on different evolutionary timescales to underpin the different commonly affected biological processes. Furthermore, this has allowed us to define clusters with different drug responses. In summary, this method enables us to identify the characteristics that account for the differences between tumors, suggest which drug or treatment is more likely to be effective for a given patient, and to explain why.

Poster 22

Maria Torralvo

IBE (CSIC-UPF)

"From Lemurs to Humans: A Comprehensive Look at DNA Methylation Across the Primate Tree"

Historically, one of the primary objectives of evolutionary biologists has been to understand the molecular basis underlying phenotypes. Gene expression regulation has long been hypothesized to be crucial in primate phenotypic diversity and species-specific adaptation, especially after the emergence of numerous pieces of evidence on the significant genetic similarity among phenotypically divergent species and on the effect of genetic and epigenetic differences in species evolution. A key player in this regulatory landscape is DNA methylation, particularly 5-methylcytosine (5mC), a stable epigenetic modification present in the vast majority of eukaryotes involved in the regulation of numerous biological processes, conditioning cell, tissue, and organism phenotypes. Aberrant DNA methylation has been shown to play a major role in human disease. The efforts to perform comparative studies on gene regulation in primates have been limited compared to the genomic ones. Consequently, the role of DNA methylation in primate evolution has only recently begun to be explored. Here, we present the most extensive dataset of whole genome DNA methylation patterns across the primate phylogeny to date, including a total of 132 different non-human primate species.

Poster 22

Using a multiple sequence alignment, we are able to analyze genome-wide CpG sites across key functional genomic regions, such as promoters and enhancers, greatly expanding the scope of previous array-based approaches. This research aims to shed light on the complex interplay between DNA methylation and gene regulation in primates, contributing to our understanding of primate evolution and offering potential insights into human health and disease, delivering a valuable integrative resource of great interest for the scientific community.

Poster 23

Laura Vilà Valls

IBE (CSIC-UPF)

"Whole-genome sequences uncover complex demographic and evolutionary history in modern Sudanese populations"

This study represents the first comprehensive effort to investigate the genetic history of modern Sudan from whole-genome sequencing (WGS). Given its vast diversity of populations, cultures, and languages, and the challenging desert/semi-arid climate of the Sahara and Sahelian belt, this region provides a unique opportunity to trace the origins of human genetic variation, such as the impact of demographic events and local adaptation. In this study, we report high coverage (~30X) WGS of 125 individuals from five Sudanese populations: Beja, Copts, Fulani, Fur, and Mahas, representing three different language families: Afro-Asiatic, Niger-Congo, and Nilo-Saharan. Combined with previously published complete genomes, this dataset allows for an unprecedented study of the genetic landscape of this region of the African continent. Our analysis aims to investigate the impact of significant demographic events, such as the Arabization, and the genetic origins of the populations under study. Additionally, we seek to identify signals of positive selection that could indicate local adaptation. Our initial findings reveal complex genetic structures within Sudan, pointing to different demographic histories. The inhabitants of Darfur show a nearly unique autochthonous genetic component, except for a small amount of West African gene flow, suggesting a long and lasting period of time without admixture.

Poster 23

Although Islamized after the Arabization and the spread of Islam eastward, Darfuri autochthonous populations did not significantly admix with the Arabic tribes that settled in Southwest Sudan, indicating a cultural but not genetic impact. However, genetic traces of Arabization are present in other Sudanese populations such as the Beja, the Mahas and the Copts, although the autochthonous genetic component has also persisted in smaller proportions. This study is particularly important because it sheds light on a region that has been underrepresented in human genomic studies. By examining these unique populations, we aim to fill existing gaps in our understanding of human genetic diversity and adaptation in the Sahara and Sahelian regions, providing a more comprehensive and representative portrait of human genetic history.

Poster 24

Lucas E. Wange

IBE (CSIC-UPF)

"Investigating the evolution of repressive gene regulatory elements in mammals"

Uncovering the rules of gene regulation is crucial for understanding how the diversity of tissues, cell types, and developmental stages arise from a single genome sequence. While much progress has been made in identifying promoters and enhancers in the past two decades since the sequencing of the human genome, silencers have been more difficult to detect on a genome-wide scale. Silencers are enriched in human disease variants similar to enhancers and are likely to have played an important role in the evolution of divergent phenotypes. Recent advances in silencer identification have improved their annotation in humans, however, evolutionary information and comparisons across species are still lacking to complete the picture and understand their contribution to disease. This project aims to identify silencers in primates and mammals to investigate their evolutionary turnover and validate lineage-specific gains or losses. Using an existing dataset of great apes, including humans, we identify silencers using a machine learning approach, creating the first catalog of candidate silencers in non-human primates. Employing a silencer specific massively parallel reporter assay we will functionally validate human-specific, great ape-specific, and primate-conserved silencer elements. Finally, we are generating a comprehensive regulatory profile of 21 placental mammals using fibroblast cell lines.

Poster 24

By integrating measurements of gene expression, chromatin accessibility, histone modifications, and whole genome methylation data, we will identify silencers and compare their evolutionary patterns across mammals. This research will enhance our understanding of silencer function, evolution and the role of repressive regulatory elements in generating phenotypic variation.

Poster 25

Illya Yakymenko

IMIM

"Exploring the functional consequences of SPPL2c variants within a positively selected inversion in human populations"

Genomic inversions are structural variants that contribute to evolutionary adaptations. The 17q21.31 inversion is a well-studied ~1 Mb-long human inversion that is associated to multiple phenotypic traits and has been proposed to be positively selected in European populations. Due to recombination suppression, the two inversion haplotypes (H1 and H2) accumulate around 3000 variants in perfect linkage disequilibrium, affecting the regulation of multiple coding and non-coding transcripts. However, linking individual variants to specific traits remains very challenging. In this study, we focus on the potential functional effect of changes in the single-exon gene SPPL2C, which is located within the inverted region and encodes a 684 amino acid aspartyl protease. Notably, SPPL2C harbors seven missense mutations between the two orientations that may alter the function of the protein. Particularly, the substitution of a highly conserved arginine at position 461 with a proline (R461P) has been classified as potentially damaging by several tools (CADD score: 23.3; SIFT score: 0.01; AlphaMissense score: 0.851). We generated molecular dynamics simulations of both SPPL2c variants with one of its substrates, VAPB, embedded in the endoplasmic reticulum membrane.

Poster 25

The R461P change apparently induces an allosteric conformational modification affecting the interaction of crucial residues of the catalytic center (D386 and D448) with VAPB. Specifically, VAPB forms a hydrogen bond with D386 during 20.8% of the simulated time in the R461 variant and with D448 for 29.0% of the time in the P461 variant. Interestingly, the other catalytic residue loses any interaction with VAPB in both cases. Thus, we hypothesize that the variants associated to the 17q21.31 inversion alter the catalytic activity of SPPL2c and, consequently, play a role in the regulation of the SNARE complex. This could be related to the protective effect against neurodegeneration of the H2 haplotype and experimental validation of the findings is currently underway.

Poster 26

Matteo Zambon

CRG

"Benchmarking models of human gastrulation through analyses of gene regulatory networks and differentiation trajectories"

Gastrulation phase is a pivotal part of mammalian embryo development. In spite of the biological and clinical relevance, its study on humans is extremely challenging, both due to technical and ethical reasons. In vitro models as human embryonic organoids are currently considered the best option to overcome such issues. Several different models of early human development have been recently published, while one crucial question remains their resemblance with in vivo embryos in terms of emerging cell types and gene expression dynamics in differentiating cell lineages. In order to get insights on how these in vitro models differentiate and to compare them with in vivo references through single-cell RNA sequencing, two complementary computational approaches are established, leveraging the inference of gene regulatory networks and the inference of gene expression patterns along the mesoderm lineage. The established procedures allowed the comparison of a human embryo, mouse embryos and several human embryonic organoids at corresponding stages of development, highlighting the resemblance across the systems in activity of transcription's master regulators in the inspected cell types while also identifying genes with distinct expression patterns during in vitro differentiation.