DR. PEDRO J. BALLESTER

INSERM-tenured group leader

Group Leader, Cancer Research Centre of Marseille (CRCM), France (10/2014 – date)

- A*MIDEX Excellence Chair laureate in *international competition* (two-year seed funding ending in 01/2017 for foreign scientists to start a group in France).
- Awarded an INSERM-tenured CR1 position in *international competition* (8% success rate in 2014).
- Awarded *HDR* (Habilitation à Diriger des Recherches) degree by the *Faculty of Medicine of Aix-Marseille University (AMU)* on March 2016 (thesis submitted on September 2015).
- Published *21 peer-reviewed papers* with this new affiliation (*16 as corresponding author*) as in 04/2018 with *5 papers more in review* (*4 as corresponding author*).
- Founded *my first research group*, CRCM's Computational Biology and Drug Design team, *funded by external grants except from one PhD scholarship*. 13 members since 10/2014: 1 group leader (tenured), 4 postdocs, 4 PhD students, 1 postdoc visitor (3 months) and 2 master project students.

Staff Scientist, EMBL-European Bioinformatics Institute, Cambridge, UK (1/2010 – 6/2014)

- Awarded a 4-year Methodology Research Fellowship from the Medical Research Council (MRC) in *UK-wide competition* to fund this position (22% success rate) hosted by Janet Thornton's group.
- Awarded a Junior Research Fellowship (JRF) at Wolfson College Cambridge in *university-wide competition with postdocs from any field of study* (~10% success rate).
- Published *17 peer-reviewed papers* (8 as *first author* and *14 as corresponding author*) on new computational methods for structure-based drug design and cancer pharmacogenomics.

Postdoc, Centre for Molecular Informatics, **Cambridge University, UK** (1/2009 – 12/2009)

- Designed and conducted a *pioneering study* demonstrating that nonlinear machine learning provides substantially better structure-based prediction of protein-ligand binding.
- Published this study in 2010 as *first and co-corresponding author* in Bioinformatics, *the highest ranked journal at the Mathematical & Computational Biology* JCR category of that year. It is now an *ISI highly cited paper* (top1% of the papers published in this category).

Postdoc, Center for Computational Drug Discovery, **Oxford University, UK** (10/2005 – 12/2008)

- Awarded a Junior Research Fellowship (JRF) at St Cross College Oxford in *university-wide competition with postdocs from any field of study* (~10% success rate).
- *Merit award for exceptional research contribution* to Oxford University's Dept. Chemistry.
- Published *9 peer-reviewed papers* (*8 as first and corresponding author*) on novel computational methods for ligand-based drug design *following a change of discipline*.

PhD Candidate, Computational Geophysics, Imperial College London, UK (10/2001 – 07/2005)

- *PhD awarded on 31/07/2005* by Imperial College, University of London, UK. PhD while working as postgraduate research assistant on an EPSRC grant awarded to Dr Jonathan Carter (supervisor).
- Published *9 peer-reviewed papers* (*7 as first and corresponding author*) on the development and application of multi-modal optimisation methods for geophysical inverse problems.

MSc Student, Department of Mathematics, Kings College London, UK (09/2000 – 09/2001)

- MSc in "Information Processing and Neural Networks*" awarded with Distinction* by King's College London (Distinction is the highest award level for a master in the UK).

Scientific Software Developer, CSIC, Spain (10/1997 – 09/2001)

- Awarded a scholarship to fund one-year taught MSc in the UK by the Spanish Sa Nostra Foundation (value €18,000; 8 scholarships were awarded to Spanish postgraduate students from any discipline).
- Working in IT support, including scientific software development but not research. Thus, this is effectively a *three-year gap period* (research career started with PhD studies in late 2001).

First Degree Student, Dept. of Physics, University of La Laguna, Spain (10/1992 – 07/1997)

- 5-year *licenciatura* in Physics with Astrophysics.

RESEARCH GRANTS

Awarded as Principal Investigator (country; net funding granted to my group)

- 11/2017 – 10/2019 ANR Tremplin-ERC (France; €130,000; Ref: ANR-17-ERC2-0003-01).
- 12/2016 – 11/2019 IPC PhD Scholarship for Ms Alexandra Bomane (France; 3 years full funding).
- 02/2015 – 01/2017 A*MIDEX Excellence Chair (France; €235,000; Ref: ANR-11-IDEX-0001-02).
- 10/2014 – 09/2016 INSERM CR1 Starting Allowance (France; €40,000; Ref: BBG/2015).
- 07/2010 – 06/2014 MRC Methodology Research Fellowship (UK; £400,905; Ref: G0902106).
- 06/2008 – 08/2008 Pfizer's travel grant to fund two visits for a total of 10 weeks to their Groton Laboratories in CT, USA (US; $15,000; Ref: R08279/CN001) – hosts: Wan Lau & Brian Bronk.

Awarded as Co-Investigator (country; net funding granted to my group)

- 2016 Canceropôle PACA (France; €83,000 ; http://canceropole-paca.com/). PI: Prof Norvert Vey.
- 2016 HEC PhD Scholarship for Ms Fahmida Ahmad (Pakistan; 3 years full funding).
- 2016 CONACyT PhD Scholarship for Ms Elva Novoa (Mexico; 3 years full funding).
- 2016 USTH PhD Scholarship for Ms Linh Nguyen (Vietnam; 3 years full funding).
- 2013 EMBLEM TDF grant (Germany; €55,000; http://www.embl-em.de/). PI: Dr Maja Koehn.
- 2011 National Research Plan (Spain; €55,400, Ref: TEC2011-23113). PI: Dr Jose L. Rossello.
- 2008 Oxford University's BtG scheme (UK; £10,841; Ref: RC/7). PI: Prof David Gavaghan.

SUPERVISION OF GRADUATE STUDENTS & POSTDOCTORAL SCIENTISTS (excl. MSc thesis)

- Supervision of **4 postdocs**: Dr Cuong Dang (02/2015–04/2017), Dr Antonio Peon (02/2015–01/2017), Dr Hongjian Li (10/2015–12/2015) and Dr Stefan Naulaerts (02/2017–07/2018).
- Supervision of **4 PhD students**: Ms Linh Nguyen (01/2016–12/2018), Ms Elva Novoa (11/2016–03/2017), Ms Fahmida Ahmad (11/2016–04/2017) and Ms Alexandra Bomane (12/2016–11/2019).

COMMISSIONS OF TRUST (Selected)

- 2018 – date *Referee* (funding proposals), Israel Science Foundation (ISF), Israel.
- 2018 – date *Referee* (funding proposals), European Commission's Horizon 2020, EU.
- 2018 – date *Advisory Editorial Board Member* of Computational and Mathematical Biophysics.
- 2017 – date *Referee* (funding proposals), The Swiss National Science Foundation, Switzerland.
- 2017 – date *Editorial Board Member* of Scientific Reports (a Nature Research journal).
- 2016 – date *Referee* (funding proposals), Netherlands Org. for Scientific Research, Netherlands.
- 2015 – date *Editorial Board Member* of Frontiers in Chemistry (Review Editor for the Medicinal and Pharmaceutical Chemistry specialty).
- 2015 – date *Referee* (funding proposals), National Agency for Research (ANR), France.
- 2014 – date *Referee* (funding proposals), National Agency for Research (ANEP), Spain.
- 2014 – date *Referee* (funding proposals), Biotechnology and Biological Sciences Research Council (BBSRC), UK.
- 2014 – date *Faculty member*, Cancer Research Centre of Marseille (INSERM U1068, CNRS UMR7258, AMU UM105, Institute Paoli-Calmettes), France.
- 2012 – 2014 *Member of the Faculty Committee for Academic Admissions* (Membership and Fellowship Committee), Wolfson College, Cambridge University, UK.
- 2011 – 2014 *Faculty member* (Governing Body Fellow), Wolfson College, Cambridge University, UK.
- 2011 – 2014 *Thesis Advisory Committee member* for Ms Birgit Hoeger (EMBL-Heidelberg, Germany) and Mr Michael Menden (EMBL-EBI, UK).
- 2011 – 2013 *Referee* (funding proposals), National Scientific Research Council, Romania.
- 2012 – 2012 *Referee* (funding proposals), National Science Foundation, Czech Republic.

OTHER CONTRIBUTIONS TO RESEARCH COMMUNITY (Selected)

As Reviewer for Journals: 184 reviews for 60 journals certified by https://publons.com/author/975063/ (Sentinels of Science award for being a top reviewer between Oct-2015 and Sep-2016).

As Reviewer for International Conferences (Selected): International Work-Conference on Bioinformatics and Biomedical Engineering (IWBBIO 2015-16), 8th and 9th IAPR International Conference on Pattern Recognition in Bioinformatics (PRIB 2013-14), 10th and 11th International Meeting on Computational Intelligence Methods for Bioinformatics and Biostatistics (CIBB 2013-14),

ECCB European Conference on Computational Biology (2012, 2014), ACM Genetic and Evolutionary Computation Conference (GECCO 2007-15), IEEE Congress on Evolutionary Computation (2005, 2007, 2009-11), IEEE Symposium Series on Computational Intelligence (2007, 2009), IEEE World Congress on Computational Intelligence (2006, 2008).

As Co-organiser of Conferences: *11th International Meeting on Computational Intelligence Methods for Bioinformatics and Biostatistics* (2014, The Computer Laboratory, Cambridge University, UK), *Tools and Strategies to Find Chemical Probes for Your Protein-The Role of Computer-Aided Drug Discovery* (2013, Biochemical Society and Royal Society of Chemistry workshop, London, UK).

Early achievements track-record

55 papers in the period 2003-2017 (50 published, 5 in review).

22 as first author (40%).

43 as single or joint corresponding author (78%).

h-index = 19 if only considering peer-reviewed papers in which I am either first or corresponding author (Source: Google Scholar).

NB: For each paper, my author position is underlined and corresponding authors in **bold** font.
First research group started in February 2015.

SELECTED PEER-REVIEWED PAPERS (none with PhD supervisor):

Nguyen, L., Naulaerts, S., Bomane, A., Bruna, A., Ghislat, G., **Ballester, P.J.** (2018) "Machine learning models to predict in vivo drug response via optimal dimensionality reduction of tumour molecular profiles" (In Review). https://www.biorxiv.org/content/early/2018/03/06/277772
(large-scale analysis of PDX drug-sensitivity and molecular profiles of tumours leading to the identification of highly predictive machine-learning models for three drugs employing concise gene lists – novel methods).

Naulaerts, S., Dang, C., **Ballester, P.J.** (2017) "Precision and recall oncology: combining multiple gene mutations for improved identification of drug-sensitive tumours". *Oncotarget* 8 (57), 97025.
(using the same genomic and drug sensitivity data, we show that machine-learning markers are able to identify a higher proportion of sensitive tumours than single-gene markers in 118 of the 127 analysed drugs).

Peon, A., Naulaerts, S., **Ballester, P.J.** (2017) "Predicting the reliability of drug-target interaction predictions with maximum coverage of target space" *Scientific Reports* 7, 3820
(a first-in-kind *in silico* tool to predict the protein targets of small organic molecules, such as hits from phenotypic screening campaigns, along with the reliabilities of such predictions).

Wójcikowski, M., **Ballester, P.J.**, **Siedlecki, P.** (2017) "Performance of machine-learning scoring functions in structure-based virtual screening". *Scientific Reports* 7, 46710. https://www.nature.com/articles/srep46710
(this study shows that machine-learning scoring functions can achieve outstanding performance at structure-based virtual screening if these scoring functions are trained with large volumes of synthetic negative data).

Li, H., Leung, K.-S., Wong, M.-H., **Ballester, P.J.** (2016) "USR-VS: a web server for large-scale prospective virtual screening using ultrafast shape recognition techniques". *Nucleic Acid Research* 44 (W1): W436-W441. (journal ranked 20th of 289 journals in the JCR Biochemistry & Molecular Biology category).

Peon, A., Dang, C., **Ballester, P.J.** (2016) "How reliable are ligand-centric methods for target fishing?" *Frontiers in Chemistry* 4:15. http://journal.frontiersin.org/article/10.3389/fchem.2016.00015
(a critical analysis of the reliability of *in silico* tools with maximum coverage of the proteome for predicting the protein targets of small organic molecules such as hits from phenotypic screening campaigns).

Ain, Q.U., Aleksandrova, A., Roessler, F.D., **Ballester, P.J.** (2015) "Machine-learning scoring functions to improve structure-based binding affinity prediction and virtual screening". *WIREs Computational Molecular Science* 5, 405–424.
(invited to write a first review on machine-learning scoring functions for molecular docking; journal currently ranking 1st out of 57 journals in the Mathematical & Computational Biology ISI JCR category).

Ballester, P.J., Schreyer, A., Blundell, T.L. (2014) "Does a more precise chemical description of protein-ligand complexes lead to more accurate prediction of binding affinity?". *Journal of Chemical Information and Modeling* 54, 944–955.
(2nd most read article in 2014 at one of the reference journals in the area of Chemoinformatics).

Menden, M., Iorio, F., Garnett, M., McDermott, U., Benes, C., **Ballester, P.J.**, **Saez-Rodriguez, J.** (2013) "Machine learning prediction of cancer cell sensitivity to drugs based on genomic and chemical properties". *PLOS ONE* 8, e61318.
(>19,000 views since publication, featured in PLOS Collections; this study integrates for the first time the chemistry of drug molecules with the genomic features of cells in modelling cancer pharmacogenomics).

Ballester, P.J., Mangold, M., Howard, N.I., Marchese-Robinson, R.L., Abell, C., Blumberger, J. and Mitchell, J.B.O. (2012) "Hierarchical virtual screening for the discovery of new molecular scaffolds in antibacterial hit identification". *Journal of the Royal Society Interface* 9:77, 3196-3207.
(this journal ranked 5th out of 56 in the Multidisciplinary Sciences category of the 2012 JCR; collaboration with experimental colleagues, reporting a prospective application of our computational methods).

Ballester, P.J. and **Mitchell, J.B.O.** (2010) "A machine learning approach to predicting protein-ligand binding affinity with applications to molecular docking". *Bioinformatics* 26:9, 1169-1175.
(this journal ranked 1st out of 47 journals in the Mathematical & Computational Biology category of the 2010 JCR; this study is the first in demonstrating the key role of the adopted regression model in the prediction of protein-ligand binding strength; 210 citations so far – Web of Science highly-cited paper (top 1%).

Ballester, P.J., Westwood, I., Laurieri, N., Sim, E., Richards, W.G. (2010) "Prospective Virtual Screening with Ultrafast Shape Recognition: The Identification of Novel Inhibitors of Arylamine N-acetyltransferases". *Journal of the Royal Society Interface* 7:43, 335-342.
(this journal ranked 4th out of 59 journals in the Multidisciplinary Sciences category of the 2010 JCR; collaboration with experimental colleagues, reports a prospective application of our computational methods).

PATENTS, COMMERCIALISATION AND CONSULTING

- Ballester, P.J. (2012) US Patent 8244483 "Shape Recognition Methods and Systems for Searching Molecular Databases" (http://www.google.com/patents/US8244483).
- ISIS Innovation commercialisation project No 2932 – Molecular Shape Recognition. Various confidential licensing and consulting contracts. Non-confidential information can be requested to the project manager Dr Jamie Ferguson (jamie.ferguson@isis.ox.ac.uk).

AWARDS, FELLOWSHIPS AND SCHOLARSHIPS

2017 *Tremplin-ERC* laureate awarded by the ANR (French Agency for National Research).
2016 *HDR* (Habilitation à Diriger des Recherches) degree awarded by the *Faculty of Medicine* of AMU.
2015 *Excellence Chair* awarded by Aix-Marseille Initiative of Excellence in international competition.
2014 CR1 *tenured position* awarded by INSERM in international competition.
2011 Junior Research *Fellowship* awarded by Wolfson College, Cambridge University, UK.
2010 Methodology Research *Fellowship* awarded by the MRC in UK-wide competition.
2007 Research *merit award* from the Department of Chemistry, Oxford University, UK.
2007 Junior Research *Fellowship* awarded by St Cross College, Oxford University, UK.
2000 *Scholarship* to fund one-year taught MSc in the UK awarded by the Spanish Sa Nostra Foundation (value €18,000; 8 scholarships were awarded to Spanish postgraduate students from any discipline).

SELECTED INVITED TALKS

In International Scientific Meetings (out of 32 talks; web link underlined)

2018 (Basel, Switzerland):BioData World *Congress*
2018 (Toronto, Canada): *Workshop* on the Mathematics of Drug Design/Discovery, The Fields Institute
2018 (London, UK): World Precision Medicine *Congress*
2018 (Cardiff, UK): Joint UK-QSAR and MGMS *meeting*
2017 (Cambridge, UK): BioData World *Congress*
2016 (B.A., Argentina): 4th ISCB Latin America *Bioinformatics* conference
2016 (Paris, France): 3rd Institut Pasteur International Network Symposium on *Biomarkers*
2016 (Q.Roo, Mexico): 6th SolBio conference on *Bioinf. & Comput. Biology for Innovative Genomics*
2015 (Shanghai, China): CSHA/AACR *Big Data, Computation and Systems Biology in Cancer* conference
2015 (Berlin, Germany): 2nd Yandex *Machine Learning: Prospects and Applications* conference
2015 (Berlin, Germany): Pharma Exabyte: *Shaping the Next Generation Pharmaceutical R&D with BigData*
2015 (Granada, Spain): 3rd International Conference on *Bioinformatics & Biomedical Engineering*
2013 (London, UK): Tools and Strategies to Find Chemical Probes for Your Protein – The Role of *Computer-Aided Drug Discovery* (Biochem. Soc. and R. Soc. Chemistry workshop)
2013 (Berlin, Germany): Intelligent Systems for Molecular Biology (ISMB) / European Conference on *Computational Biology* (ECCB) 3DSig

2012 (Tokyo, Japan): 7th IAPR International Conference on *Pattern Recognition in Bioinformatics*
2007 (Sheffield, UK): 4th Joint Sheffield Conference on *Chemoinformatics*
2006 (Canberra, Australia): Australian Academy of Science White Conference on *Mastering the Data Explosion* in the Earth and Environmental Sciences
2004 (Seattle, USA): 13th ACM *Genetic and Evolutionary Computation* Conference

In Academic Institutions (out of 19 talks; invitation from named scientist)
2015 (Shanghai, China): CAS–MPG Institute for Computational Biology (Prof Andrew Teschendorff).
2013 (Cambridge, UK): Computational Biology Institute, Cambridge University (Dr Christophe Dessimoz).
2012 (Canterbury, UK): School of Computing, University of Kent (Dr Dominique Chu).
2010 (Heidelberg, Germany): EMBL–Heidelberg (Dr Maja Koehn).
2009 (Hamburg, Germany): EMBL–Hamburg (Prof Matthias Wilmanns).
2008 (Oxford, UK): Computational Biology group, Oxford University (Prof David Gavaghan).
2008 (Cambridge, UK): Structural Bioinformatics group, Cambridge University (Prof Tom Blundell).

In Industry (out of 8 talks; invitation from named scientist)
2013 (Cambridge, UK): Computational Chemistry group, Astex Pharmaceuticals (Dr Marcel Verdonk).
2010 (Madrid, Spain): HTS group, GlaxoSmithKline (Dr Julio Martin).
2008 (Groton, USA): Computational Science Seminars, Pfizer (Dr Wan Lau).
2007 (Harlow, UK): Computational Chemistry Europe, GlaxoSmithKline (Dr Colin Edge).