TITLE OF THE PROJECT (ACRONYM): *In vivo* determinants of 3D enhancer-promoter communication *(3DCOMM)*

BACKGROUND INFORMATION

In this project, we aim to understand how enhancer elements can find their correct target genes and activate them with spatial and temporal precision. As such, the goal of the project is to improve and refine current models of 3D gene regulation, being framed within the context of basic research. This fits well within the selected modality of the current call, which covers non-oriented research projects.

Our research group is located at the Centro Andaluz de Biología del Desarrollo (CABD) in Sevilla and is currently funded by european (ERC, EMBO) and spanish (AEI) funding agencies. We have a particular interest in understanding mechanisms of 3D gene regulation *in vivo*. During the past years, we have actively developed CRISPR genome editing techniques and tetraploid aggregation methods. This unique combination of technologies provides great versatility in generating transgenic mouse models. Importantly, it substantially reduces the economic, personnel and time requirements for transgenic work, while better complying with the 3R principles of animal research. We employ this experimental setting to systematically investigate gene regulation at selected loci and subsequently extrapolate our findings to a genome-wide scale. Thus, while we have a marked profile for basic research, our *in vivo* setup provides a deeper understanding of gene regulation within developmental and disease contexts.

Developmental gene regulation in a 3D space

During development, a temporal and spatial control of gene expression is key for the appropriate differentiation of cellular types, tissues and organs. In metazoans, transcription is largely controlled by *cis*-regulatory elements (CREs), such as enhancers (*1*). These elements, often located at large genomic distances from their promoters, act as binding platforms for transcription factors and define where and when genes are expressed. Multiple CREs can be associated to the same gene, exerting redundant functions that provide robustness to developmental gene expression (*2–4*). Advances in genomics approaches have dramatically accelerated our understanding on the mechanisms of transcriptional control (*5*, *6*). Such methods allowed us to transition from models that merely described the transcriptional relations between genes into complex *cis*-regulatory networks that provide a better contextualization of developmental processes. Yet, while the identification and prediction of CREs has become a routine, it still remains difficult to elucidate the specificity of a target gene(s) for a given enhancer (*7–9*).

To activate gene expression, CREs enter into spatial proximity with their cognate gene promoters (*10*) (although the “degree ”of proximity is a matter of debate (*11*)). The emergence of high-resolution microscopy and of chromatin conformation capture (3C) methodologies (*12*, *13*), provided means to detect and quantify the interactions between genomic regions within the nucleus. The application of these methods, in particular of a genome-wide high-throughput variant (Hi-C), revealed that metazoan genomes are organized into topologically associating domains (TADs) (**Fig. 1**). TADs represent large genomic regions, ranging from a few hundred kilobases to megabases, that are defined by boundaries. Loci that occupy the same TADs generally interact with high frequency, while interactions across boundaries are largely restricted (*14–16*). Consequently, genes within the same TADs display a higher degree of coexpression than those located in separate domains. Furthermore, interactions between CREs and target genes rarely occur across TAD boundaries (*17*, *18*). Despite some exceptions (*19*), TADs are conserved to a large extent among cell types and species, highlighting the relevance of this organizational unit (*20*, *21*). As such, TADs exert important constraints in genome evolution (*22*), by favoring rearrangement to occur at boundaries (*23*, *24*) or, as we recently showed, through the disappearance of entire TAD units after the second round of whole genome duplication in vertebrates (*25*). We and others have demonstrated the functional relevance of TADs and their boundaries, by showing that their disruption, by means of genomic rearrangements, can cause human pathologies that include congenital malformations and cancer (*3*, *26–30*). Furthermore, we recently highlighted the role of TAD rearrangements as evolutionary drivers of phenotypical innovation, as exemplified by the formation of ovotestes in female moles (*31*) or the expansion of fin morphology in skates (*25*).

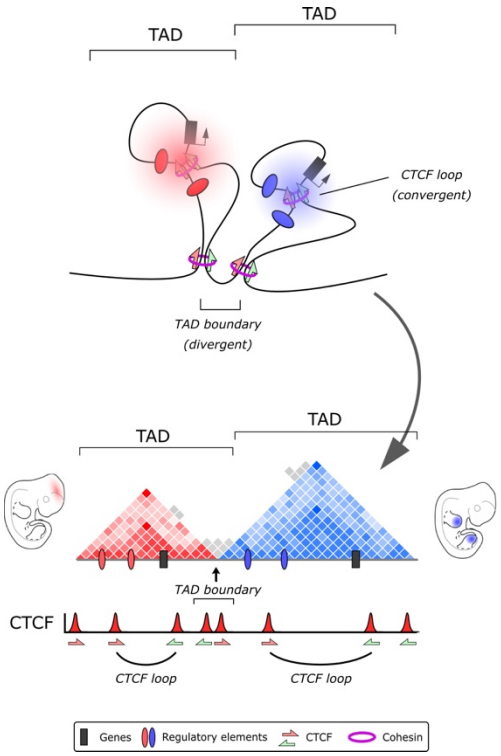

Figure 1. A CTCF code in mammalian genomes. *Convergently-oriented CTCFs can block the cohesin complex, promoting interaction between distant genomic regions. At TAD boundaries, divergently-oriented CTCF-clusters limit interactions between adjacent TADs.*

In vertebrates, most TADs are formed by a mechanism of loop extrusion for which the transcriptional repressor CCCTC-binding factor (CTCF) plays an important role (*32*). CTCF encodes 11 zinc fingers that binds to a conserved, non-palindromic 20 bp consensus motif (*33*, *34*). The effects of CTCF in gene regulation were initially linked to its ability to function as an insulator, capable of isolating promoters from the effects of nearby regulatory elements (*35*) However, recent work in vertebrate systems highlights a much more extensive role for CTCF in genome regulation. CTCF is found at most vertebrate TAD boundaries (*14*, *36*), mediating their insulator capacity (**Fig. 1**). Equally importantly, CTCF appears to mediate long-range interactions within TADs, by forming chromatin loops that connect distant regulatory elements and promoters (*36*, *37*). The ability of CTCF to mediate chromatin interactions is intimately linked to cohesin, a ring-shaped protein complex that can extrude chromatin and bring distal loci into physical proximity. This process, known as loop extrusion, can be interrupted when the cohesin complex encounters a genomic region bound by CTCF, with its DNA-binding motif displaying a convergent orientation to the direction of extrusion (*38*, *39*). As such, convergent motifs are preferentially found at pairs of CTCF-looping sites (*36*, *40*, *41*), and chromatin interactions can be redirected upon inversion of CTCF motifs (*42*). At TAD boundaries, CTCF-binding-sites (CBS) often cluster together and with divergent orientations, forming a characteristic corner loop at the top of the upstream and downstream TADs that can be observed in Hi-C maps. Such clustered CBS play redundant functions that confer robustness to the insulator function of these regions, by limiting inter-TAD interactions and promoting intra-TAD contacts (*43*). As a result of the loop extrusion process, genomic regions engaged in chromatin loops display increased interaction. Importantly, this property has been shown to regulate promoter activation by distal enhancers (*44*). The depletion of CTCF or of cohesin

subunits leads a general loss of TADs and loops (*45–47*). Yet, certain degree of 3D chromatin organization remains and even becomes more apparent, mainly resulting from the interactions between loci that display a similar transcriptional status (A/B compartmentalization). This demonstrates the existence of additional mechanisms of 3D chromatin organization. Regarding this, a subset of TAD boundaries are CTCF-independent and can be associated to housekeeping (*14*, *48*), highly transcribed genes (*49*, *50*) or transposable elements (*51–53*) highlighting the role of transcription in shaping genome organization. Yet, some studies suggest that is not transcription itself, but rather the binding of RNA polymerase what may cause the formation of TAD boundaries (*54–57*). Nevertheless, the act of transcription can still affect 3D chromatin organization (*58–60*). Therefore, vertebrate genomes are spatially organized as a result of two independent mechanisms, one related to transcription and another based in CTCF- and cohesin-mediated loop extrusion.

Why an enhancer activates a specific gene (and ignore others)?

Even if the discovery of TADs represents a compelling framework to understand how enhancers find their target gene(s), it only provides a partial explanation for this phenomenon. While TADs display increased levels of transcriptional coregulation, it is not uncommon that these domains contain multiple genes with distinct patterns of expression. Actually, enhancer elements generally target specific genes, while ignoring others that might be located even at shorter distances in the linear genome (*61*). To date, the molecular principles that govern this intriguing property are still an area of intense research.

3D chromatin organization is certainly a fundamental layer contributing to enhancer-promoter specificity. As described in the previous section, structural variants that disrupt TAD boundary regions can expand the genomic range of action of enhancer elements, potentiall interacting with non-canonical genes to cause their misexpression (*27*, *62*). These effects derive from the disruption of insulating factors, like CTCF, that act at boundary regions. Yet, insulating factors can be also found at other genomic locations. For example, a recent study showed that CTCF is bound to more than 60% gene promoters in different mouse tissues (*44*), and mutations that abrogate such binding can result in decreased interactions with the target gene and activation of nearby genes, as shown at particular loci (*63*) and at a genome-wide scale (*44*). As also noted previously, promoters of highly transcribed regions have been also associated with boundaries (*14*, *48–53*), suggesting an additional role as insulating elements.

Another explanation for enhancer-promoter specificity is the compatibility between these elements. Classical examples for this phenomenon can be observed at the mouse *Mrf4/Myf5* (*64*) or at the *Shh/Lmbr1* loci (*65*), in which tissue-specific enhancers regulate genes selectively but fail in activating others. The development of Massive Parallel Reporter Assays (MPRA), such as STARR-seq, allowed the investigation of the enhancer-promoter compatibility phenomenon at a genome-wide scale (*66*, *67*). These studies revealed that promoters and enhancers can be categorized into two broad classes, developmental and housekeeping, each showing regulatory preferences for elements belonging to their own class. A systematic analysis with episomal plasmids yielded similar results, even finding different compatibilities within developmental genes (*68*). Such additional differences may be explained by additional genomic features of developmental promoters, such as CpG islands, which can act as transcriptional booster for certain classes of enhancers (*69*). In addition, genes that are compatible for specific enhancers can compete for these elements. For example, a gain-of-function mutation at the alpha-globin locus creates a novel promoter that blocks the canonical activation of globin genes in an orientation-dependent manner, causing alpha-thalasemia (*70*). Loss-of-function mutations can lead to opposite effects, inactivating promoters and causing a redirection of enhancer activity to other nearby genes (*64*, *71*). Deciphering the complex principles of enhancer-promoter specificity is essential to fully understand gene regulation.

The Epha4 locus: a paradigm to study gene regulation

The *Eph receptor A4* (*Epha4*) locus represents a powerful model that has helped us to uncover important principles of gene regulation. *Epha4* is a gene that is prominently expressed during limb development and which inactivation does not cause major morphological consequences, besides some defects in hindlimb innervation (*72*). This gene is located in a 2 megabases (Mb) TAD, where it interacts with a downstream cluster of limb enhancers that control its expression (*73*). Structural variants associated to severe limb phenotypes in human

patients were mapped to this TAD, for which alterations in *EPHA4* regulation could not offer a plausible disease mechanism. In particular, a 1.6Mb deletion disrupted the telomeric boundary region (EP boundary) of this TAD, as well as the *EPHA4* gene itself, and was associated with brachydactyly (shorter anterior digits) in human patients (*73*). By developing a method to generate large structural variants in mouse models (*74*), we demonstrated that this deletion led to a fusion of the *Epha4* and its adjacent TAD (*28*, *73*). This fusion led to the ectopic limb expression of the *Paired box 3* (*Pax3*) gene by a novel interaction with *Epha4* limb enhancers, that are otherwise isolated from this gene in healthy controls (**Fig. 2**). In contrast, a smaller deletion that did not include the EP boundary region was sufficient to preserve the partition between the *Epha4* and *Pax3* TADs, and did not have further consequence in gene expression and phenotypes. Importantly, similar TAD-rearranging mechanisms can also affect at the centromeric boundary of the *Epha4* TAD, leading to the appearance of syndactyly (soft fusion of fingers) of polydactyly (additional fingers) phenotypes, through the ectopic activation of the *Wnt family member 6* (*Wnt6*) and *Indian hedgehog* (*Ihh*) genes, respectively (*28*, *73*). This denotes the promiscuity of *Epha4* limb enhancers in inducing the expression of other developmental genes besides their canonical target. Therefore, by studying the molecular origin of certain limb malformation in human patients, in combination with mouse models, we provided the first evidence that disruptions of TAD organization could cause congenital diseases.

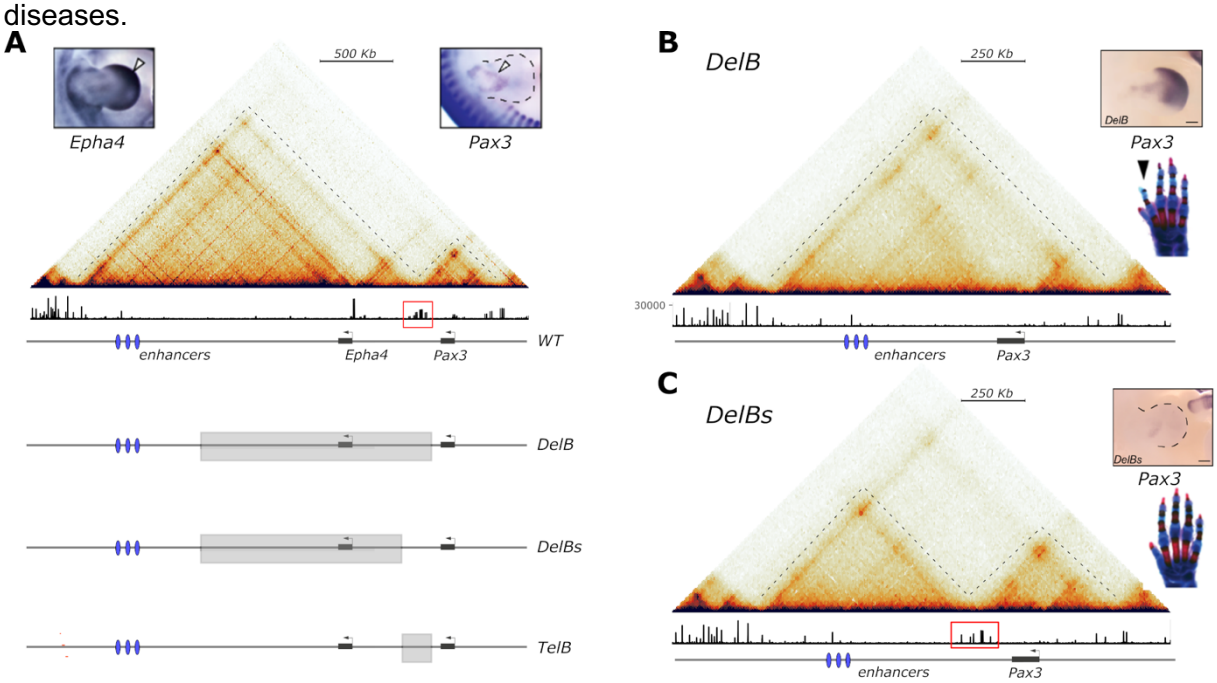

Figure 2. 3D gene regulation at the Epha4 locus and effects of deletions. *(A) cHi-C maps for the Epha4 locus at E11.5 limb development. Note the partition of the locus in two TADs (dashed line) separated by the EP boundary region (red rectangle). Above, whole mount in situ hybridization (WISH) for Epha4 and Pax3. Note that Epha4 is expressed in the distal limb, while Pax3 in the proximal limb (arrows). Below, CTCF ChIP-seq track and schematic of the locus. Bottom, schematics of deletions performed at the locus (TelB deletion is further described in Fig. 3. Gray rectangle marks the deleted regions (B) Data from DelB background. Note the fusion of TADs (dashed line), distal misexpression of Pax3 resulting from interaction with Epha4 enhancers and brachydactyly in skeletal preparations from adult limbs. (C) Data from DelBs background. The deletion is similar to DelB, but not affecting the EP boundary region (red rectangle). Note the preserved partition of TADs (dashed line), normal expression of Pax3 and normal phenotype in skeletal preparations.*

In subsequent studies, we explored the regulatory logic of CTCF-associated boundaries. We particularly focused on the EP boundary, which is a prototypical boundary composed of 6 CBS displaying the divergent orientation that is characteristic of these regions (*43*). By performing individual and combined deletions of CBS, we discover that these sites act in a cooperative and redundant manner to constitute insulation at boundary regions, as also described in *in vitro* screening studies (*75*). Importantly, we observed that not all CBS are functional equivalent, reporting the existence of a novel class of sites that can anchor simultaneous chromatin loops in convergent and non-convergent directions. Such sites, which

we named bidirectional CBS, display a robust functionality and stall cohesin complexes persistently in the same orientation as their binding motifs. Such stalling can block additional complexes coming from the opposite direction, thus creating an additional paired loop in a non-canonical orientation. This mechanism, which we termed as loop interference, was previously predicted by the loop extrusion model (*38*) and observed *in vitro* (*76*). Yet, our study constituted the first validation for this mechanism *in vivo*. Our study also discovered a quantitative relation between the degree of disruption of EP boundary function, the misexpression of *Pax3* and the severity of brachydactyly in mutant mice. This makes the *Epha4* locus an attractive model to decipher the relation between 3D chromatin interactions, gene expression and developmental phenotypes.

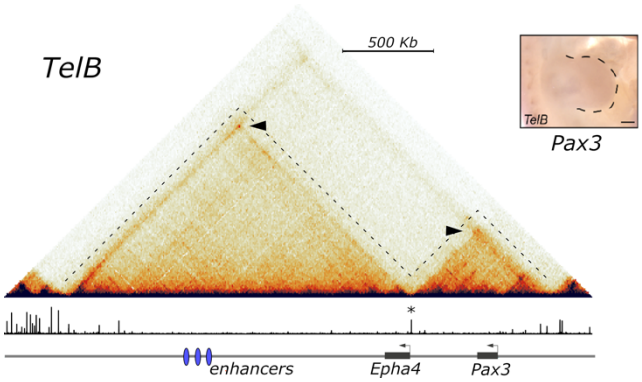

Figure 3. 3D gene regulation in the TelB background. *cHi-C maps for the Epha4 locus at E11.5 limb development. Above, WISH for Pax3. Below, CTCF ChIP-seq track and schematic of the locus. Note the preserved partition of TADs and normal expression of Pax3. The boundary separating the Epha4 and Pax3 TADs is associated to the Epha4 promoter and a CBS (asterisk) that anchor loops in both directions (arrows)*

In preliminary experiments that led to this project proposal, we investigate the *in vivo* effects of deleting of the EP boundary alone (**Fig. 3**). Strikingly, transgenic mice carrying this 150 kilobases (kb) deletion (*TelB* background) did not show any phenotype, nor any evidence of *Pax3* misexpression. This is in stark contrast with our previous findings that demonstrate the sufficiency of the EP boundary region to isolate the *Epha4* and *Pax3* TADs (*DelB* and *DelBs* backgrounds) (*28*, *73*). However, our new results demonstrate that the EP boundary it is also not strictly necessary to segregate both TADs and that additional sequences at the locus can replace this function. These results are consistent with observations at the *Sox9* and *Shh* locus, where deletions of boundaries do not lead to major changes in 3D gene regulation (*77*, *78*). Capture Hi-C experiments (cHi-C) in mouse mutant limbs served us to identify the novel boundary region, which is located near with the *Epha4* promoter, raising the hypothesis that this transcriptional unit is involved in the segregation of the *Epha4* and *Pax3* TADs. Furthermore, a downstream-oriented bidirectional CBS near the *Epha4* promoter also serves as an anchor for chromatin loops that project in simultaneous directions. In our previous study at the EP boundary, we observed that even the presence of two robust bidirectional CBS was insufficient to prevent the aberrant interaction between TADs and *Pax3* misexpression. These results, together with the previously reported role of CBS and promoters in shaping 3D chromatin organization, led us to postulate that both the *Epha4* and its associated CBS are both functional components of the boundary observed in the *TelB* background.

As such, these observations raise several fundamental questions that we aim to systematically address at the *Epha4* locus, such as:

How do CBS and promoters cooperate to insulate 3D regulatory domains?
How does CBS or promoter characteristics modulate transcriptional outputs?
How do genes compete for the same enhancers and influence each other?

Our hypothesis is that CBS and promoters form a regulatory interplay that ensures correct gene expression during development. On the one hand, these two types of elements may ensure that enhancers enter into physical proximity with the appropriate gene. On the other hand, they may restrict aberrant interactions with non-target genes. Our deep understanding of the *Epha4* locus regulation and our newly generated mutant *TelB* background represent a unique opportunity to gain insights in the determinants of enhancer-promoter communication and their impact during the development of a living organism.

OBJETIVES

The **main goal** of this project is **to systematically dissect the functional interplay between CBS and promoters in guiding enhancers to their target genes**. We will leverage on the *Epha4* locus, a developmental model that has provided important advances in our understanding of gene expression, during development and in disease contexts (see also section 3.5). Importantly, our experimental design has the advantage that it can measure the effects of genomic perturbations *in vivo* at three different level (3D chromatin interactions, gene expression and phenotypes). This setup allows to establish causal relationships at a quantitative level while we decouple two important aspects of gene regulation: the interaction of enhancers with a cognate promoter (*Epha4*) and their isolation from non-target but responsive genes (*Pax3*). Our capability of capturing variations in spatial transcription patterns and in phenotypes serves to better frame our results within the context of developmental processes. By generating a series of 11 mutations on the mouse *TelB* background and measuring gene expression, 3D chromatin interactions and phenotypes, we will:

- ***Aim1:* Dissect the relative contribution of CBS and promoters in insulator function.** We will investigate the functional boundary near the *Epha4* promoter in the *TelB* background, which effectively segregates the *Epha4* and *Pax3* TADs. By performing combined and individual deletions and quantifying 3D chromatin organization and *Pax3* misexpression *in vivo*, we will determine the relative contribution of the *Epha4* promoter and of its associated bidirectional CBS in insulator function.
- ***Aim 2:* Investigate how CBS and promoter functionality modulates selective gene activation**. We will replace the bidirectional CBS at the *Epha4* promoter by alternative configurations, such as unidirectional CBS or altered orientations. In parallel, we will replace the *Epha4* promoter by classes of promoter types, such as housekeeping or non-CpG associated. We will determine *in vivo* how these replacements alter the 3D organization of the locus, both affecting the canonical expression of *Epha4* and inducing the ectopic activation of *Pax3*.
- ***Aim 3:* Determine how changes in canonical transcription relate to non-target gene activation.** In a series of selected mutants, we will measure gene expression at single-molecule resolution to quantify how changes in *Epha4* and *Pax3* relate to each other, in individual cells and in different limb regions. We will further evaluate how expression changes correlate to the appearance of two distinct phenotypes: hopping gait (related to *Epha4* insufficiency) and brachydactyly (related to *Pax3* misexpression).

TRAINING PROGRAM

The PhD student will be enrolled in the Biotechnology doctoral program from the University Pablo de Olavide. This is a very multidisciplinary program that covers basic and applied research and that helps developing training competences. Those competences include experimental design, critical thinking or communication skills, among others. A personalized career development plan will be devised with the PhD student, supervised by the PI. This will include a yearly meeting with a Thesis Advisory Committee, which will be formed two members from CABD, and an external member. The student will also have an "Expectations and Feedback Meeting" with the PI every six months that will be more focused in personal and professional development. The student will also benefit from the environment of our research group. I have been leading an international and multicultural laboratory and I plan to maintain a similar environment at the CABD. Weekly lab meetings are specifically designed to improve presentation skills, creative thinking and data interpretation. The student will also be integrated to the research environment of the CABD, in which wet and computational researchers are exposed to a wide range of research topics in development, disease and evolution. The development of computational skills for experimental scientists is encouraged by finding appropriate mentors and the assistance to training courses. The student will also take part in journal clubs, as well as seminars that occur regularly at the institute, and that feature national and international scientist of high profile. The student will also benefit from the integration of

our group in international consortia and in the EMBO YIP network, all offering opportunities to receive training and to present her/his project in an international environment. Furthermore, short visits to international groups will be planned, so that the student can experience the research culture of other countries. For example, a will include a visit to the laboratory of Blanche Capel in Duke University (USA) is already agreed withing the context of this project. The assistance to international congresses and meeting to present the project will be encouraged at all phases of the project. Outreach activities like Open Day activities, Science Cafes will be also planned.

REFERENCES

1. H. K. Long *et al.*, *Cell*. **167**, 1170–1187 (2016).
2. E. Z. Kvon *et al.*, *Nature Reviews Genetics*, 1–13 (2021).
3. A. J. Will *et al.*, *Nat. Genet.* **49**, 1539–1545 (2017).
4. M. Osterwalder *et al.*, *Nature*. **554**, 239–243 (2018).
5. S. Bonn *et al.*, *Nat Genet*. **44**, 148–156 (2012).
6. A. Rada-Iglesias *et al.*, *Nature*. **470**, 279–283 (2011).
7. Q. Cao *et al.*, *Nat. Genet.* **49**, 1428–1436 (2017).
8. D. Hafez *et al.*, *Genome Biol.* **18**, 199 (2017).
9. R. P. Zinzen *et al.*, *Nature*. **462**, 65–70 (2009).
10. E. E. M. Furlong *et al.*, *Science*. **361**, 1341–1345 (2018).
11. N. S. Benabdallah *et al.*, *Mol Cell*. **76**, 473-484.e7 (2019).
12. J. Dekker *et al.*, *Science*. **295**, 1306–1311 (2002).
13. E. de Wit *et al.*, *Genes Dev.* **26**, 11–24 (2012).
14. J. R. Dixon *et al.*, *Nature*. **485**, 376–380 (2012).
15. T. Sexton *et al.*, *Cell*. **148**, 458–472 (2012).
16. E. P. Nora *et al.*, *Nature*. **485**, 381–385 (2012).
17. O. Symmons *et al.*, *Genome Res.* **24**, 390–400 (2014).
18. F. Le Dily *et al.*, *Genes Dev.* **28**, 2151–2162 (2014).
19. W. Winick-Ng *et al.*, *Nature*. **599**, 684–691 (2021).
20. R. D. Acemel *et al.*, *Wiley Interdiscip Rev Dev Biol*. **6** (2017).
21. M. Franke *et al.*, *Current Opinion in Cell Biology*. **55**, 24–29 (2018).
22. R. D. Acemel *et al.*, *Curr Opin Genet Dev*. **78**, 102019 (2023).
23. D. Li *et al.*, *BMC Biology*. **20**, 99 (2022).
24. J. Krefting *et al.*, *BMC Biology*. **16**, 87 (2018).
25. F. Marlétaz *et al.*, *BioRxiv* (2022), p. 2022.03.21.485123.
26. D. G. Lupiáñez *et al.*, *Trends Genet.* **32**, 225–237 (2016).
27. M. Spielmann *et al.*, *Nat. Rev. Genet.* (2018).
28. S. Bianco *et al.*, *Nat. Genet.* (2018).
29. W. A. Flavahan *et al.*, *Nature*. **529**, 110–114 (2016).
30. D. Hnisz *et al.*, *Science*. **351**, 1454–1458 (2016).
31. F. M. Real *et al.*, *Science*. **370**, 208–214 (2020).
32. R. Zhang *et al.*, *Exp. Cell Res.* **294**, 86–93 (2004).
33. D. Martin *et al.*, *Nat. Struct. Mol. Biol.* **18**, 708–714 (2011).
34. K. Van Bortle *et al.*, *Genome Res.* **22**, 2176–2187 (2012).
35. P. E. Szabó *et al.*, *Mol Cell Biol*. **24**, 4791–4800 (2004).
36. S. S. P. Rao *et al.*, *Cell*. **159**, 1665–1680 (2014).
37. G. Ren *et al.*, *Mol. Cell*. **67**, 1049-1058.e6 (2017).
38. G. Fudenberg *et al.*, *Cell Rep*. **15**, 2038–2049 (2016).
39. A. L. Sanborn *et al.*, *PNAS*. **112**, E6456–E6465 (2015).
40. C. Gómez-Marín *et al.*, *PNAS*. **112**, 7542–7547 (2015).
41. M. Vietri Rudan *et al.*, *Cell Rep*. **10**, 1297–1309 (2015).
42. Y. Guo *et al.*, *Cell*. **162**, 900–910 (2015).
43. C. Anania *et al.*, *Nat Genet*. **54**, 1026–1036 (2022).
44. N. Kubo *et al.*, *Nat Struct Mol Biol*. **28**, 152–161 (2021).
45. E. P. Nora *et al.*, *Cell*. **169**, 930-944.e22 (2017).
46. W. Schwarzer *et al.*, *Nature*. **551**, 51–56 (2017).
47. G. Wutz *et al.*, *The EMBO Journal*. **36**, 3573–3599 (2017).
48. S. Hong *et al.*, *Nucleic Acids Res.* **45**, 10403–10414 (2017).
49. A. R. Barutcu *et al.*, *Nat Commun*. **9**, 1444 (2018).
50. C. L. Harrold *et al.*, *BioRxiv* (2020), p. 2020.07.01.182089.
51. V. V. Lunyak *et al.*, *Science*. **317**, 248–251 (2007).
52. J. Wang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **112**, E4428-4437 (2015).
53. Y. Zhang *et al.*, *Nat Genet*. **51**, 1380–1388 (2019).
54. Z. Du *et al.*, *Nature*. **547**, 232–235 (2017).
55. M. J. Rowley *et al.*, *Mol. Cell*. **67**, 837-852.e7 (2017).
56. B. Bonev *et al.*, *Cell*. **171**, 557-572.e24 (2017).
57. C. B. Hug *et al.*, *Cell*. **169**, 216-228.e19 (2017).
58. S. Heinz *et al.*, *Cell*. **174**, 1522-1536.e22 (2018).
59. G. A. Busslinger *et al.*, *Nature*. **544**, 503–507 (2017).

60. J. M. Brown *et al.*, *Nat Commun*. **9**, 3849 (2018).
61. C. C. Galouzis *et al.*, *Curr Opin Cell Biol*. **75**, 102065 (2022).
62. C. Anania *et al.*, *Brief Funct Genomics*. **19**, 128–138 (2020).
63. S. Oh *et al.*, *Nature*. **595**, 735–740 (2021).
64. J. J. Carvajal *et al.*, *Genes Dev*. **22**, 265–276 (2008).
65. L. A. Lettice *et al.*, *Hum. Mol. Genet.* **12**, 1725–1735 (2003).
66. D. T. Bergman *et al.*, *Nature*. **607**, 176–184 (2022).
67. M. A. Zabidi *et al.*, *Nature*. **518**, 556–559 (2015).
68. M. Martinez-Ara *et al.*, *Molecular Cell*. **82**, 2519-2531.e6 (2022).
69. T. Pachano *et al.*, *Nat Genet*. **53**, 1036–1049 (2021).
70. Y. K. Bozhilov *et al.*, *Nat Commun*. **12**, 3806 (2021).
71. S. W. Cho *et al.*, *Cell*. **173**, 1398-1412.e22 (2018).
72. F. Helmbacher *et al.*, *Development*. **127**, 3313–3324 (2000).
73. D. G. Lupiáñez *et al.*, *Cell*. **161**, 1012–1025 (2015).
74. K. Kraft *et al.*, *Cell Rep* (2015).
75. H. Huang *et al.*, *Nat Genet*. **53**, 1064–1074 (2021).
76. A. Allahyar *et al.*, *Nat Genet*. **50**, 1151–1160 (2018).
77. A. Despang *et al.*, *Nature Genetics*. **51**, 1263–1271 (2019).
78. I. Williamson *et al.*, *Development*. **146** (2019).
79. J. Artus *et al.*, *Methods Mol. Biol.* **693**, 37–56 (2011).
80. K. Kruse *et al.*, *Genome Biology*. **21**, 303 (2020).
81. E. de Wit *et al.*, *Mol. Cell*. **60**, 676–684 (2015).
82. ENCODE Project Consortium, *Nature*. **489**, 57–74 (2012).
83. H. M. T. Choi *et al.*, *Development*. **145**, dev165753 (2018).
84. M. Arzt *et al.*, *Frontiers in Computer Science*. **4** (2022).
85. I. Mota-Gómez *et al.*, *BioRxiv* (2022), p. 2022.11.18.516861.

CURRICULUM VITAE OF THE IP

Part A. PERSONAL INFORMATION *(*) Mandatory*

First name	Darío Jesús		
Family name	Lupiáñez García	ID number	75888504G
Gender (*)	Male	Birth date (dd/mm/yyyy)	04/11/1981
e-mail	Dario.Lupianez@mdc-berlin.de	URL Web: https://lupilab.wordpress.com/	
Open Researcher and Contributor ID (ORCID)(*)	0000-0002-3165-036X		

A.1. Current position

Position	Group Leader	Initial date	01.09.2017
Institution	Max Delbrück Center for Molecular Medicine		
Department/Center	Berlin Institute for Medical Systems Biology (BIMSB)		
Country	Germany	Teleph. number	+49-30-9406-3085
Key words	Epigenetics, 3D chromatin, functional genomics, development		

A.2. Previous positions

Period	Position/Institution/Country/Interruption cause
2015-2017	Project group leader/Max Planck Institute for Molecular Genetics/Germany
2012-2015	Postdoctoral researcher/Max Planck Institute for Molecular Genetics/Germany
2006-2011	PhD student/University of Granada/Spain
2004-2005	Scientific Assistant/University of Granada/Spain

A.3. Education

PhD, Licensed, Graduate	University/Country	Year
Ph.D., Genetics. Doctor Europeus	University of Granada/Spain	2011
Diploma in Advanced Studies	University of Granada/Spain	2007
Bachelor in Pharmacy	University of Granada/Spain	2004

Part B. CV SUMMARY *(max. 5000 characters, including spaces)*

My research focuses on the mechanisms controlling 3D gene regulation and their influence on developmental phenotypes.

I performed my PhD at the University of Granada (Rafael Jiménez group), investigating the development of ovotestes in female moles. During this period, we establish the mole as a research model and discovered a series of molecular events associated to the sexual features of this species. Specifically, we found that female moles display an extreme heterochrony in female meiosis (Zurita et al., *Sex Dev*, 2007). We also discover that mole ovotestes develop a profuse vascular system that facilitate the export of androgens to the mole body (Lupiáñez et al., *J Exp Zool Part B*., 2012). Further, we identified that mole ovotestis development is associated to alterations on the spatiotemporal expression pattern of key regulators of sex determination (Carmona et al., *Int J Dev Biol*., 2009 and *J Exp Zool Part B*., 2009).

Then, I performed a postdoc at the Max Planck Institute in Berlin (Stefan Mundlos group). We developed a novel method to model large structural variants in transgenic mice (Kraft et al., *Cell Reports*, 2015). With this methodology, we were the first to report that disruptions on 3D chromatin organization can cause congenital disease (Lupiáñez et al., *Cell*, 2015; Will et al, *Nat Genet*. 2017; Kragesteen et al, *Nat Genet*. 2018). We established a conceptual framework for interpreting the effects of structural variants on 3D gene regulation (Lupiáñez et al, *Trends Genet*. 2016; Spielmann et al., *Nat Rev Genet*. 2018), as well as developing predictive computational approaches (Bianco et al., *Nat Genet*. 2018).

In 2017, I started my lab at the Berlin Institute for Medical Systems Biology/Max Delbrück Center for Molecular Medicine (BIMSB-MDC). Here, I studied how alterations in 3D gene regulation may underlie the emergence of phenotypical traits during evolution. An example is our recent study on the mole genome and on the molecular causes of intersexuality (Real et al., *Science*, 2020). We developed a novel phylogenomic analytical framework to identify species-specific structural variants that affect 3D chromatin organization. We subsequently identified mole-specific variants associated to ovotestis development and validated them in transgenic mice. We also applied this approach to investigate the molecular origin of wing-like fins in skates, revealing an involvement of the PCP pathway in anterior fin elongation (Marlétaz et al. *BioRxiv*, 2022). Another prominent area of research relates to the characterization of the molecular mechanisms that drive 3D chromatin organization *in vivo.* One example is our recent molecular characterization of TAD boundary elements, revealing their cooperative nature (Anania et al., *Nat Genet*. 2022).

I have received several fellowships and awards, including an ERC Consolidator Grant, the EMBO Young Investigator or the ESHG Young Investigator Award. I have also secured several grants (total funding > 3 million €). I have 27 publications and 3 preprints, including first or senior author in high-impact journals like *Cell*, *Science* or *Nature Genetics* with a total of 3,3557 citations (Google Scholar). I have been invited lecturer at more than 50 international research centres and conferences. I have experience as main organiser of international conferences (8 in total), including 2 editions of the EMBO workshop “The evolution of animal genomes”. I have supervised 19 students (7 PhD). I have reviewed for several funding agencies, including the *European Research Consortium* (ERC), and journals including *Science, Cell* or *Nature Genetics*. I am associated editor of *Science Advances* since 2020 and member of the editorial board of *“Nucleus”* since 2022.

Recently, I obtained a tenured scientist position (científico titular CSIC) at the Centro Andaluz de Biología del Desarrollo (CABD-CSIC/UPO/JA) in Seville, which I will join in July 2023.

Part C. RELEVANT MERITS *(sorted by typology, chronologically inverted, 10 each typology)*

C.1. Publications *(* equally contribution, AC corresponding author)*

- Mota-Gómez I*, Rodríguez JA*, Dupont S, *et al.,* Capel B* (AC), Marti-Renom M* (AC), **Lupiáñez DG* (AC)**. (13/13) 2022. *Sex-determining 3D regulatory hubs revealed by genome spatial auto-correlation analysis.* BioRxiv doi: https://doi.org/10.1101/2022.11.18.516861. (*Manuscript under revision in Nature*)
- Marlétaz F.* (AC), de la Calle-Mustienes E.*, Acemel R.D.*, *et al.,* Tena J.J* (AC), **Lupiáñez D.G* (AC)**, Rokhsar D.S* (AC), Gómez-Skarmeta J.L*. (29/31). 2022. *The little skate genome and the evolutionary emergence of wing-like fin appendages.* BioRxiv. doi: https://doi.org/10.1101/2022.03.21.485123. (*Acceptance pending minor editorial changes in Nature*)

- Anania, C*., Acemel, R.D*., Jedamzick, J., *et al.,* **Lupiáñez D.G (AC)**. (10/10). 2022. *In vivo dissection of a clustered-CTCF domain boundary reveals developmental principles of regulatory insulation.* Nature Genetics 54(7):1026-1036
- Real, F.M., Haas, S.A., Franchini, P., *et al.,* Mundlos, S.*(AC), **Lupiáñez D.G**.* **(AC).** (29/29). 2020. *The mole genome reveals regulatory rearrangements associated with adaptive intersexuality.* Science 70(6513):208-214.
- Kragesteen B.K., Spielmann M., Paliou C., *et al.,* **Lupiáñez D.G.,** *et al.,* Mundlos S., Andrey G. (17/24) 2018. *Dynamic 3D chromatin architecture contributes to enhancer specificity and limb morphogenesis.* Nature Genetics 50(10):1463-1473.
- Bianco S.*, **Lupiáñez D.G.*,** Chiariello A.M.*, *et al.,* Mundlos S (AC), Nicodemi M (AC). (2/12). 2018. Polymer physics predicts the effects of structural variants on chromatin architecture. Nature Genetics 50(5):662-667.
- Will, A.J., Cova, G., Osterwalder, M., *et al.,* **Lupiáñez, D.G.*(AC),** Mundlos, S*(AC). (12/13). 2017. *Composition and dosage of a multipartite enhancer cluster control developmental expression of Indian hedgehog.* Nature Genetics 49(10):1539-1545.
- **Lupiáñez, D.G.,** Kraft, K., Heinrich, V., *et al.,* Mundlos, S (AC). (1/22). 2015. *Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions.* Cell 161(5):1012-1025.
- Kraft, K., Geuer, S., Will, A.J., Chan, W.L., Paliou, C., Borschiwer, M., Harabula, I., Wittler, L., Franke, M., Ibrahim, D.M., Kragesteen, B.K., Spielmann, M., Mundlos, S., **Lupiáñez, D.G.* (AC),** Andrey, G* (AC). (14/15). 2015. *Deletions, Inversions, Duplications: Engineering of Structural Variants using CRISPR/Cas in Mice.* Cell Reports 10 (5), 833-839.
- **Lupiáñez, D.G.,** Real, F.M., Dadhich, R.K., Carmona, F.D., Burgos, M., Barrionuevo, F.J., Jiménez, R. (1/7). 2012. *Pattern and density of vascularization in mammalian testes, ovaries and ovotestes.* Journal of Experimental Zoology Part B: Molecular and Developmental Evolution 318(3):170-81.

C.2. Congresses

- 2022. EMBO Workshop: The Many Faces of Cancer Evolution, Rimini, Italy. **Invited speaker**
- 2021. EMBO Workshop: The Evolution of Animal Genomes. **Invited speaker**.
- 2021. EMBO Workshop: Enhanceropathies: Understanding Enhancer Function to Understand Human Disease, Santander, Spain. **Invited speaker**.
- 2021. European Human Genetics Conference. Virtual. **Invited speaker**.
- 2021. Spanish Society of Genetics (SEG) Congress. Virtual. **Invited speaker**.
- 2019. Workshop “From Genes to Organisms: Transcriptional Control in Development and Disease”, Baeza, Spain. **Invited speaker**.
- 2019. EMBO Workshop “Beyond the standard: Non-model vertebrates in biomedicine”, Berlin, Germany. **Invited speaker**.
- 2018. Spanish Society for Biochemistry and Molecular Biology Conference (SEBBM), Santander, Spain. **Invited speaker**
- 2017. Joint Congress 2017 (SEG- SEBC- SEBD), Gijon, Spain. **Invited speaker**
- 2016. American Society of Human Genetics 2016 Annual Meeting, Vancouver, Canada. **Invited Talk**

C.3. Research projects

- Ref: 101045439. *The impact of 3D regulatory landscapes on the evolution of developmental programs (3D-REVOLUTION).* **ERC** Consolidator Grant, call 2021. 2023-2028. 1,998.218 €. **Principal investigator**.
- Ref: 422856854. Evolution of 3D chromatin architecture: The role of CTCF across taxa. Priority Programm “Spatial Genome Architecture in Development and Disease” - SPP 2202 from Deutsche Forschungsgemeinschaft (**DFG**, Germany). Call 2018. 2019-2022; 235.150 €. **Principal investigator.**
- Ref: 355312821. *The regulatory landscape of mammalian sex determination.* Project grant from the International Research Training Group “Dissecting and Reengineering the

Regulatory Genome" - IRTG 2403 from Deutsche Forschungsgemeinschaft (**DFG**, Germany). Call 2018. 2019-2022; 247.909 €. **Principal investigator.**

- Ref: Helmholtz ERC Recognition Award 2019. *"From cells to organs: Understanding gene regulatory dynamics in development and evolution (CELL2ORGAN)"*. **Helmholtz-Gemeinschaft**, Germany. Call 2018. 2019-2020. 200.000 €. **Principal investigator.**
- Ref: 331208046. *Molecular Analysis of Topologically Associating Domains (TADs).* Project Grant from Deutsche Forschungsgemeinschaft (**DFG**, Germany). Call 2016. 2017-2021; 233.116 €. **Principal investigator.**
- Ref: 282307777. *Genomic Biology of Limb and Gonad Development in the Spanish Mole (Talpa occidentalis)*. Deutsche Forschungsgemeinschaft (**DFG**, Alemania). PI Stefan Mundlos. Call 2015. 2016-2020. 300.000 €. **Researcher**.
- Ref: CGL2015-67108-P. *Mecanismos de regresión y activación testicular en mamíferos con reproducción estacional: control genético y susceptibilidad a cambios ambientales.* Ministerio de Economía y Competitividad (**AEI**, España). PI Rafael Jimenez. Call 2015. 2016-2019. 74.400 € + 1 contrato predoctoral. **Researcher**.
- Ref: CGL2008-00828/BOS. *Elementos genéticos ambientales que controlan los cambios gonadales en mamíferos con reproducción estacional.* Ministerio de Ciencia e Innovación (**AEI**, España). PI Rafael Jiménez Medina. Call 2018. 2009-2011. 199.000€. **Researcher**.
- Ref: CVI2057. *Implicación de miRNAs en el control de genes de determinación sexual.* Proyectos de Excelencia de la **Junta de Andalucía**. Consejería de Innovación Ciencia y Empresa. IP Rafael Jiménez Medina. Call 2006. 2007-2009. 201.636,00 € + 1 contrato predoctoral. **Researcher**.
- Ref: CGL2004-00863/BOS. *Desarrollo gonadal y reversión sexual en Tálpidos.* Ministerio de Ciencia e Innovación (**AEI**, España). IP Rafael Jiménez Medina. Call 2004. 2005-2007. 55.600,00 € + 1 contrato predoctoral. **PhD student associated to the project**.

C.4. Honors, awards and fellowships

- **2022 -** ERC Consolidator Grant 2021. European Research Council.
- **2022 -** Research Premium 2022. Max Delbrück Center for Molecular Medicine, Germany.
- **2022 -** EMBO Young Investigator. European Molecular Biology Organization.
- **2019 -** Research Premium 2019. Max Delbrück Center for Molecular Medicine, Germany.
- **2015 -** Young Investigator Award for Outstanding Science. European Human Genetics Conference 2015. Glasgow, Scotland, United Kingdom.
- **2015 -** GFH presentation award 2015. 26th Annual Meeting of the German Society of Human Genetics. Graz, Austria.
- **2013 -** Postdoctoral fellowship. Fundación Alfonso Martín Escudero. Spain.
- **2005 -** FPI fellowship. Ministerio de Ciencia e Innovación Project (PI Rafael Jimenez).

C.5. Mentee´s predoctoral and postdoctoral fellowships

- 2021-2023. Dr. Christian Feregrino. EMBO long-term fellowship.
- 2021-2023. Dr. Rafael Acemel. EMBO long-term fellowship.
- 2020-2022. Dr. Alicia Hurtado.Fundación Alfonso Martín Escudero. Spain.
- 2019-2023. Mrs. Irene Mota. IRTG 2403 PhD fellowship
- 2018-2022. Mrs. Liene Astica. NYU PhD program from MDC

C.6. Teaching and supervision

- 2021-2022. Supervision of Mrs. Henrike Scherrer. **Graduate Student.**
- Since 2018. Supervision of Mrs. Liene Astica, Vicky Chung and Irene Mota. **PhD Students.**
- 2017-2022. Supervision of Mrs. Chiara Anania. **PhD Student.**
- 2016-2017. Supervision of Mrs. Liene Astica. **Master thesis.**
- 2016-2017. Supervision of Mr. Filip Angelov and Niklas Gerdhardt. **Graduate Student.**
- 2015-2016. Supervision of Mrs. Magdalena Schindler. **Bachelor thesis.**

- 2013-2017. Co-supervision of Mrs. Anja Will. **PhD Student.**
- 2013-2016. Co-supervision of Mrs. Katerina Kraft. **PhD Student.**
- 2014. Supervision of Mrs. Marina Borschiwer. **Bachelor thesis.**
- 2007-2009. Teaching practical **courses** (Genetics), University of Granada, Spain, 120 hours.

C.7. Organization of events

- 2023 - EMBO Workshop "The Evolution of Animal Genomes" (co-organizers: Maria Ina Arnone, Hopi Hoekstra. Sevilla, Spain.
- 2021 - EMBO Workshop "The Evolution of Animal Genomes" (co-organizers: José Luis Gómez-Skarmeta, Manuel Irimia, Fernando Casares). Virtual.
- 2020 - Berlin Summer Meeting (co-organizers: Elly Tanaka, Jan Philipp Junker, Robert Zinzen, Mina Gouti, Suphansa Sawamiphak, Daniela Panakova). Virtual.
- Since 2020 - José Luis Gómez-Skarmeta Zoominar Series (co-organizer: Denis Duboule; twice per month). Virtual.
- 2019-2020 - Systems Biology Lecture Series "Emerging views in 3D Genomics" (co-organizer: Ana Pombo). Berlin, Germany.
- 2019-2020 - Berlin Single-Cell Club (co-organizers: Nils Blüthgen, David Garfield, Jan Philipp Junker, Robert Zinzen; once per month). Berlin, Germany.
- 2019 - Workshop "From Genes to Organisms: Transcriptional Control in Development and Disease" (co-organizers: Eileen Furlong, Axel Visel). Baeza, Spain.
- 2016 - Workshop "CRISPR/Cas: Tips and Tricks" at European Human Genetics Conference (ESHG) 2016 (co-organizer: Malte Spielmann). Barcelona, Spain

C.8. Outreach activities

- **Divulgation publication.** Francisca M. Real, Stefan Mundlos, **Darío G. Lupiáñez**. 2021. *Equalizing strength among sexes: generalized intersexuality in female moles*. The Science Breaker. DOI: https://doi.org/10.25250/thescbr.brk591.
- **Television documentary.** Medizinforschung im Aufbruch - Ist Crowdfunding und vernetztes Forschen die Zukunft? 2018. 3sat, Germany.
- **Divulgation talks.** Structural Variation of the Human Genome: Order and Disorder. Future Medicine 2017. Berlin, Germany.
- **Crowdfunding project.** The LilBubome - Sequencing LilBub's Magical Genome. 2015. https://experiment.com/lilbub Daniel M. Ibrahim, Uschi Symmons, and **Darío G Lupiáñez**.
- **Divulgation events.** Lange Nacht der Wissenschaften (LNdW), years 2016, 2017, 2019.