


## RESEARCH ARTICLE

# Disrupted protein interaction dynamics in a genetic neurodevelopmental disorder revealed by structural bioinformatics and genetic code expansion

Valerio Marino<sup>1</sup> | Wanchana Phromkrasae<sup>2</sup> | Michele Bertacchi<sup>2</sup> |  
 Paul Cassini<sup>2</sup> | Krittalak Chakrabandhu<sup>2</sup> | Daniele Dell'Orco<sup>1</sup> |  
 Michèle Studer<sup>2</sup> 

<sup>1</sup>Department of Neurosciences, Biomedicine and Movement Sciences, Section of Biological Chemistry, University of Verona, Verona, Italy

<sup>2</sup>University Côte d'Azur, CNRS, Inserm, iBV, Nice, France

## Correspondence

Michèle Studer, iBV—Institut de Biologie Valrose, Univ. Côte d'Azur, Centre de Biochimie; UFR Sciences, Parc Valrose, 28 avenue Valrose, 06108 Nice Cedex 2, France.

Email: [michele.studer@unice.fr](mailto:michele.studer@unice.fr)

## Funding information

Agence Nationale de la Recherche, Grant/Award Number: ANR-15-IDEX-01; Fondation pour la Recherche Médicale, Grant/Award Number: #EQU202003010222; Fondation de France, Grant/Award Number: #00123416; ERA-NET Neuron, Grant/Award Number: ANR-21-NEU2-0003-03; Next Generation EU/Ministry of University, Grant/Award Numbers: CUP B33C22001060002, PE00000006 missione 4; Italian Ministry of Education, University and Research under the PON R&I 2014-2020 programme (FSE React-EU)

**Review Editor:** Aitziber L. Cortajarena.

## Abstract

Deciphering the structural effects of gene variants is essential for understanding the pathophysiological mechanisms of genetic diseases. Using a neurodevelopmental disorder called Bosch-Boonstra-Schaaf Optic Atrophy Syndrome (BBSOAS) as a genetic disease model, we applied structural bioinformatics and Genetic Code Expansion (GCE) strategies to assess the pathogenic impact of human NR2F1 variants and their binding with known and novel partners. While the computational analyses of the NR2F1 structure delineated the molecular basis of the impact of several variants on the isolated and complexed structures, the GCE enabled covalent and site-specific capture of transient supramolecular interactions in living cells. This revealed the variable quaternary conformations of NR2F1 variants and highlighted the disrupted interplay with dimeric partners and the newly identified co-factor, CRABP2. The disclosed consequence of the pathogenic mutations on the conformation, supramolecular interplay, and alterations in the cell cycle, viability, and sub-cellular localization of the different variants reflect the heterogeneous disease spectrum of BBSOAS and set up novel foundation for unveiling the complexity of neurodevelopmental diseases.

## KEYWORDS

BBSOAS, cellular biology, CRABP2, genetic code expansion, ligand binding domain, NR2F1, pathogenic variants, protein interactions, protein stability and affinity, structural biochemistry

Valerio Marino and Wanchana Phromkrasae have contributed equally to this study.

Krittalak Chakrabandhu, Daniele Dell'Orco, and Michèle Studer have been considered as co-last authors.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2024 The Authors. *Protein Science* published by Wiley Periodicals LLC on behalf of The Protein Society.

## 1 | INTRODUCTION

Untangling the correlation between amino acid mutations in a protein sequence and disease is crucial to understanding protein functional variation and designing effective therapeutic interventions. A mutation or truncation in the coding region of a particular gene can cause structural alterations, which may render the protein non-functional (“loss of function”). In addition, missense mutations can also lead to “gain of function” or “dominant negative” effects, resulting in functional dysregulation due to the formation of toxic aggregates or to the interference with the normal function of the wild-type (WT) protein, respectively (Dobson, 2003). Mutations located in or near key functional sites are the most likely to affect protein functions. Along with accurate information about protein structure, computational prediction of functional sites and cellular experiments can explain the effect and heterogeneity of mutations on proteins. Computational tools can also be used to calculate free energy changes associated with pathogenetic mutations, thus allowing predictions of whether the mutation will destabilize the structure of the protein, possibly affecting its function (Capriotti et al., 2004; Capriotti et al., 2005; Dell’Orco, 2009).

In this study, we focus on NR2F1 (Nuclear Receptor Subfamily 2 Group F Member 1), an evolutionary well-conserved orphan nuclear receptor acting as a strong transcriptional regulator of several genes and playing key roles during embryogenesis with a particular emphasis on the development of the central nervous system (Tocco et al., 2021). Haploinsufficiency of *NR2F1*, due mainly to *de novo* missense/nonsense mutations or whole-gene deletion of only one of the two alleles, leads to a monogenic neurodevelopmental disease, called Bosch–Boonstra–Schaaf Optic Atrophy Syndrome (BBSOAS; OMIM 615722; ORPHA 401777), an autosomal dominant genetic disorder first described in 2014 (Bosch et al., 2014). To date, BBSOAS has been diagnosed in more than 300 patients worldwide, however, new patients are reported every year, suggesting that the predicted prevalence between one in 100,000 and 250,000 people could be an underestimation (Bertacchi et al., 2022; Schaaf et al., 1993). BBSOAS symptoms are very heterogeneous in terms of both presence and severity, and include optic nerve atrophy (OA) or optic nerve hypoplasia (ONH), cortical/cerebral visual impairment (CVI), moderate to severe intellectual disability (ID), developmental delay (DD), hypotonia, seizures, speech difficulties, motor dysfunctions and autism spectrum disorder (ASD), among others. It is the peculiar combination of these diverse symptoms, particularly CVI and OA, that differentiate BBSOAS patients from those affected by other

neurodevelopmental diseases with related features (Bertacchi et al., 2022; Chen et al., 2016; Rech et al., 2020).

Although similar symptoms are shared by multiple BBSOAS patients, their degree of severity is variable, possibly depending on the location and type of the NR2F1 pathogenic variants (Bertacchi et al., 2022; Billiet et al., 2022; Chen et al., 2016; Rech et al., 2020). BBSOAS mutations are principally located in the two most conserved functional domains of the protein: the DNA-binding domain (DBD) and the ligand-binding domain (LBD). While the DBD consists of two zinc-finger domains and is responsible for the interaction with direct repeats of the consensus sequence (AGGTCA) in the promoter of target genes (Tang et al., 2015), the LBD is predicted to be necessary for protein dimerization and binding to coregulators, as suggested by the protein structure of other nuclear receptors of the same family (Rastinejad et al., 2015). Clinical investigations suggested a genotype-to-phenotype correlation, as BBSOAS patients with reduced protein dosage and functional haploinsufficiency, due to the loss of one copy of the NR2F1 gene, show a less severe clinical picture than patients with missense point mutations located in the DBD (Bertacchi et al., 2022; Bosch et al., 2014; Chen et al., 2016; Rech et al., 2020). Since NR2F1 seems to bind the DNA in the form of homodimers or heterodimers to diverse targets (Cooney et al., 1992; Klinge et al., 1997; Leng et al., 1996; Park et al., 2003), point-mutation variants might result in dominant-negative effects, in which the mutated form competes for dimerization with the WT protein or with other nuclear receptors of the same family (Tocco et al., 2021).

NR2F proteins are named “orphan” receptors since the identity of a ligand binding to the LBD domain is still elusive. Nevertheless, a functionally relevant region of the NR2F1 protein structure is the C-terminal Activation Function 2 helix (named AF2), whose active conformational state, generally obtained *via* interactions with specific ligands, allows the binding of co-factors to the LBD and ultimately controls the transcriptional regulation of target genes (Germain et al., 2006). In the specific case of the NR2F1 homolog, NR2F2, crystallographic studies have shown that the LBD is normally present in an auto-inhibited conformation, due to the binding between the AF2 helix and co-factor binding sites, and that this auto-repressed state can be reverted with a high concentration of retinoic acid (RA) (Kruse et al., 2008). Due to the high sequence identity between NR2F1 and NR2F2, it is conceivable that a similar mechanism might also exist for NR2F1. Moreover, it is still unclear whether mutations in the LBD are associated with different clinical symptoms and severity, despite the fact that a genotype–phenotype



correlation is starting to emerge (Bertacchi et al., 2022; Rech et al., 2020). Finally, little is known about the structural function of the LBD and how missense mutations or truncations within this domain specifically affect NR2F1 protein structure and, consequently, cell behavior in pathological conditions.

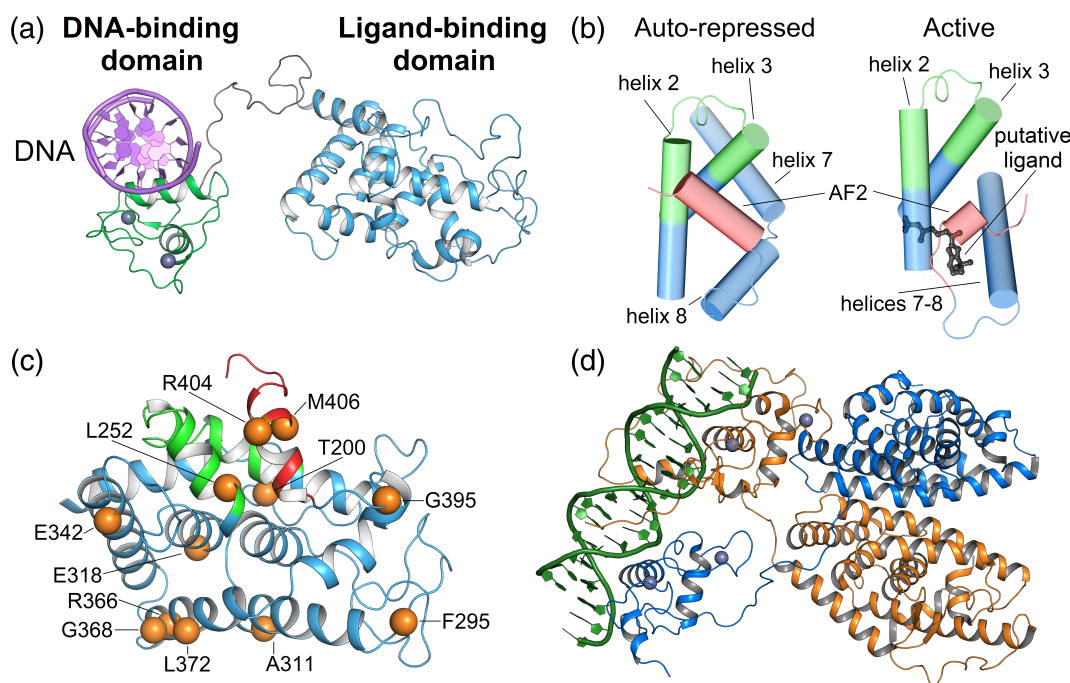
In this study, we used a multidisciplinary approach ranging from structural biochemistry to cellular biology and genetic code expansion (GCE) to specifically assess the impact of disease-associated LBD variants on the function of the NR2F1 protein. We found that some patient-specific LBD mutations show altered structural stability, impact cell proliferation and survival, and have abnormal nuclear versus cytoplasmic localization. Moreover, we found that some NR2F1 pathogenic variants display altered oligomerization both *in silico* and *in cellula* and, unexpectedly, increased propensity to form homo/heterodimers or large protein complexes. By using a GCE-enabled covalent and site-specific capturing technique, we assessed the impact of NR2F1 mutations on the dimerization and unveiled for the first time an

interaction between NR2F1 and CRABP2, which are co-expressed *in vivo* in the developing mouse brain and eyes. Together, our data shed new light on the impact of BBSOAS LBD mutations on NR2F1 activity and, more in general, on the use of structural analyses and GCE-based approaches to unravel the molecular function of nuclear receptors in both physiological and pathological conditions.

## 2 | RESULTS

### 2.1 | Molecular modeling of the NR2F1 LBD in different functional states by homology

NR2F1 is an orphan nuclear receptor belonging to the NR2F subfamily of the nuclear hormone receptor superfamily and thus shares the same structural organization as all other members of the family (Figure 1). An N-terminal disordered region is connected to the DNA-



**FIGURE 1** Structural representation of DNA-NR2F1 assembly. (a) A three-dimensional structure of NR2F1-DNA complex was modeled by the superimposition of the modeled LBD and DBD-DNA complex on the template structure provided by *AlphaFold* after removal of residues 1–83 due to unreliable predictions. Protein and DNA structure are represented as a cartoon, with the LBD shown in cyan, the DBD in green, and the dsDNA in purple. (b) Conformational changes of the AF2 helix belonging to the LBD upon ligand binding. Protein structures proximal to AF2, namely helices 2, 3, 7, and 8, are shown as blue cylindrical cartoons with AF2 helix displayed in red; the ligand is represented as black sticks with O atoms colored in red, CRS is colored in green. (c) The three-dimensional structure of the LBD is shown as a cyan cartoon with the AF2 helix shown in red and the CRS in green. The C $\alpha$  of the residues whose mutations are associated with BBSOAS are represented by orange spheres. (d) Representation of the putative interaction of NR2F1 dimer with DNA based on the homology with RXR $\alpha$ -LXR $\beta$  (PDB entry 4NQA) heterodimer. Protein structure is represented as a cartoon with RXR $\alpha$  shown in blue and LXR $\beta$  in orange, DNA strands are depicted in green.

binding domain (DBD), which is comprised of two zinc finger motifs arranged in a C4-type domain (residues 84–153) and is connected by a flexible hinge region to the all  $\alpha$ -helical ligand-binding domain (LBD, Figure 1a and Video S1). Currently, no experimental structure of the full-length human NR2F1 is available on the Protein Data Bank (PDB), and the structure predicted by *AlphaFold* is not reliable in either the N-terminal disordered region or the hinge region. Thus, by taking advantage of the high sequence identity between NR2F1 and its homolog NR2F2 (97%, Figure S1), we modeled the three-dimensional structure of the auto-repressed conformation of the NR2F1 LBD by homology using the experimentally resolved structure of the NR2F2 LBD (PDB entry 3cju) (Kruse et al., 2008).

The LBD is responsible for the dimerization of nuclear receptors *via* the dimerization interface (DI, residues 340–380 [Perlmann et al., 1996], Figure S1) as well as for the recognition and binding of ligands, coactivators, and corepressors *via* the coactivator recognition site (CRS, residues 228–253, Figure 1b) (Gampe Jr. et al., 2000; Lu et al., 2020). Accessibility of functionally important molecules to LBD is modulated by the conformational switch occurring at the C-terminal region of the protein, which involves the AF2 domain (residues 399–408, Figure 1b). Indeed, in the auto-repressed form helices 7 and 8 are separated by a loop allowing the AF2 helix to mask the CRS, while upon ligand binding, thus in the active form, the two helices fuse together and displace the AF2 helix from the CRS, which then becomes available for coregulators (Figure 1b). Similarly to the auto-repressed conformation, active NR2F1 LBD was modeled using as a template the structure of another member of the nuclear receptor superfamily, namely the Retinoid X Receptor RXR $\alpha$  (~40% sequence identity with NR2F1 LBD) in complex with 9-*cis*-retinoic acid (9cRA, PDB entry 1FM6), after removal of the coactivator peptides, as previously described in (Gampe Jr. et al., 2000; Khalil et al., 2022).

## 2.2 | Impact of BBSOAS-associated LBD mutations on protein structure

Since missense mutations in the LBD have been associated with a variety of BBSOAS symptoms (Tables 1 and S1) (Bertacchi et al., 2022; Rech et al., 2020), we predicted the effects of these mutations on the folding of both the auto-repressed and the active forms of the LBD (Figure 1c, Tables S2 and S3, and Video S2).

By comparing the calculated relative change in folding free energy ( $\Delta\Delta G_f^{\text{app}}$  values, see Section 5, Tables S2 and S3), we found that all but T200R and G368D variants

destabilize the folding of the isolated domain (both auto-repressed and active), with the largest effect exhibited by substitutions to proline (L252P, A311P, and L372P). This was not surprising, since substitutions to proline are known to disrupt the H-bond pattern required for the proper folding of the  $\alpha$ -helices on which all three residues are located. The F295L variant, localized on the long loop connecting helices 3 and 4, was predicted to exert a stabilizing effect on the active form ( $\Delta\Delta G_f^{\text{app}} = -4.60$  kcal/mol) due to a favorable hydrophobic packing within the protein core; however, an opposite effect ( $\Delta\Delta G_f^{\text{app}} = 5.1$  kcal/mol) was observed on the auto-repressed form, most probably due to the loss of the stabilizing interaction between H298 and the aromatic ring of F295, which is not present upon the F to L substitution. In addition, both G395S and G395A variants, belonging to the loop connecting helix 8 and AF2, displayed a significantly larger  $\Delta\Delta G_f^{\text{app}}$  in the auto-repressed form compared to the active one, as the conformational change pushes residue G395 outside of the protein core (Figure 1b), thus making the residue significantly more solvent-exposed and therefore more tolerant towards any potential mutation.

The structural impact of the LBD nonsense mutations, resulting in the truncation of the protein at the level of Q244 and E400, was evaluated by running exhaustive all-atom 500 ns molecular dynamics (MD) simulations (Figure 2). The analysis of the truncated structures highlighted that out of the total nine  $\alpha$ -helices constituting the LBD (Figure 2a), the E400\* variant (Figure 2b) lacked only the C-terminal AF2 helix (Figure 2a, cyan), whereas the Q244\* variant (Figure 2c) retained only helix 1 (residues 183–194; marked in yellow in Figure 2a) and helix 2 (residues 219–236; purple in Figure 2a). In terms of the structural evolution over the simulated timeframe, the C $\alpha$  Root-Mean Square Deviation (RMSD) profile calculated with respect to the equilibrated structure (Figure 2d), suggested that the E400\* truncation was overall slightly less prone to structural rearrangements over time compared to the WT form, with a ~1 Å lower RMSD in the final 50 ns of the trajectory. On the contrary, the Q244\* truncation displayed an abrupt increase in RMSD over the first 10 ns of simulation to values exceeding 6 Å, followed by a further increase to ~9.5 Å during the rest of the trajectory (Figure 2d). This strongly points towards a significantly increased plasticity of the Q244\* variant, whose lack of the largest part of secondary structure elements leads to a structural rearrangement to a different but relatively stable conformation, as shown by the comparatively smaller fluctuations in RMSD during the final part of the trajectory. Interestingly, the comparison of the RMSD of the Q244\* and E400\* truncated variants and the WT

TABLE 1 List of known LBD NR2F1 variants and clinical description of corresponding BBSOAS patients.

References	Variant (DNA)	Variant (protein)	DD	ID	Visual system defect(s) and visual deficit	EOE/S	ASD behavioral abnormalities	Hypotonia
SA13	c.1211G>A	p.Arg404His	ND	ND	ND	ND	ASD	ND
BO14	c.755T>C	p.Leu252Pro	Yes	Yes (IQ 55–65)	P/SOD; CVI; GVI	ND	ND	Yes
CH16	c.1103G>A	p.Gly368Asp	Yes	Yes (IQ ND); speech delay	No or ND	First generalized seizure at 18yo	ASD; RB; aggressive behavior	No
KA17	c.1115T>C	p.Leu372Pro	Yes, DMD	Yes (IQ ND); speech delay	OA; GVI	ND	RB; ADHD	Yes
FO19	c.603_606del	p.Arg202Thrfs*154	ND	ND	ND	ND	ND	ND
FO19	c.954G>C	p.Glu318Asp	ND	ND	ND	ND	ND	ND
FO19	c.968A>C	p.Lys323Thr	ND	ND	ND	ND	ND	ND
FO19	c.1016C>T	p.Ala339Val	ND	ND	ND	ND	ND	ND
FO19	c.1025A>G	p.Glu342Gly	ND	ND	ND	ND	ND	ND
FO19	c.1117C>T	p.Arg373*	ND	ND	ND	ND	ND	ND
FO19	c.1147_1149del	p.Ser383del	ND	ND	ND	ND	ND	ND
FO19	c.1147_1149del	p.Ser383del	ND	ND	ND	ND	ND	ND
FO19	c.1158G>T	p.Glu386Asp	ND	ND	ND	ND	ND	ND
FO19	c.1183G>T	p.Gly395Cys	ND	ND	ND	ND	ND	ND
FO19	c.1184G>A	p.Gly395Asp	ND	ND	ND	ND	ND	ND
BE20	c.729_730delinsCT	p.Gln244*	Yes	Yes	No or ND	ND	Behavioral disorders	Yes
BE20	c.967_968delAA	p.Lys23Serfs*73	Yes, DMD	Yes (speech difficulties)	OA; LVA	ND	ASD traits	No
ZO20	c.602C>A	p.Ser201*	Mild/moderate	Mild/moderate	Bilateral P/SOD; LVA	ND	ND	ND
WA20	c.1080del	p.Asn362fs*33	Apparent at 8mo	Speech delay	Severe GVI	Myoclonic atastic seizures at 2 1/2yo	ASD	ND
RE20	c.931G>C	p.Ala311Pro	Yes	Mild (FSIQ 80 below average); speech delay	P/SOD; mild GVI	Generalized Myoclonic and absence seizures	ASD	Yes

(Continues)

TABLE 1 (Continued)

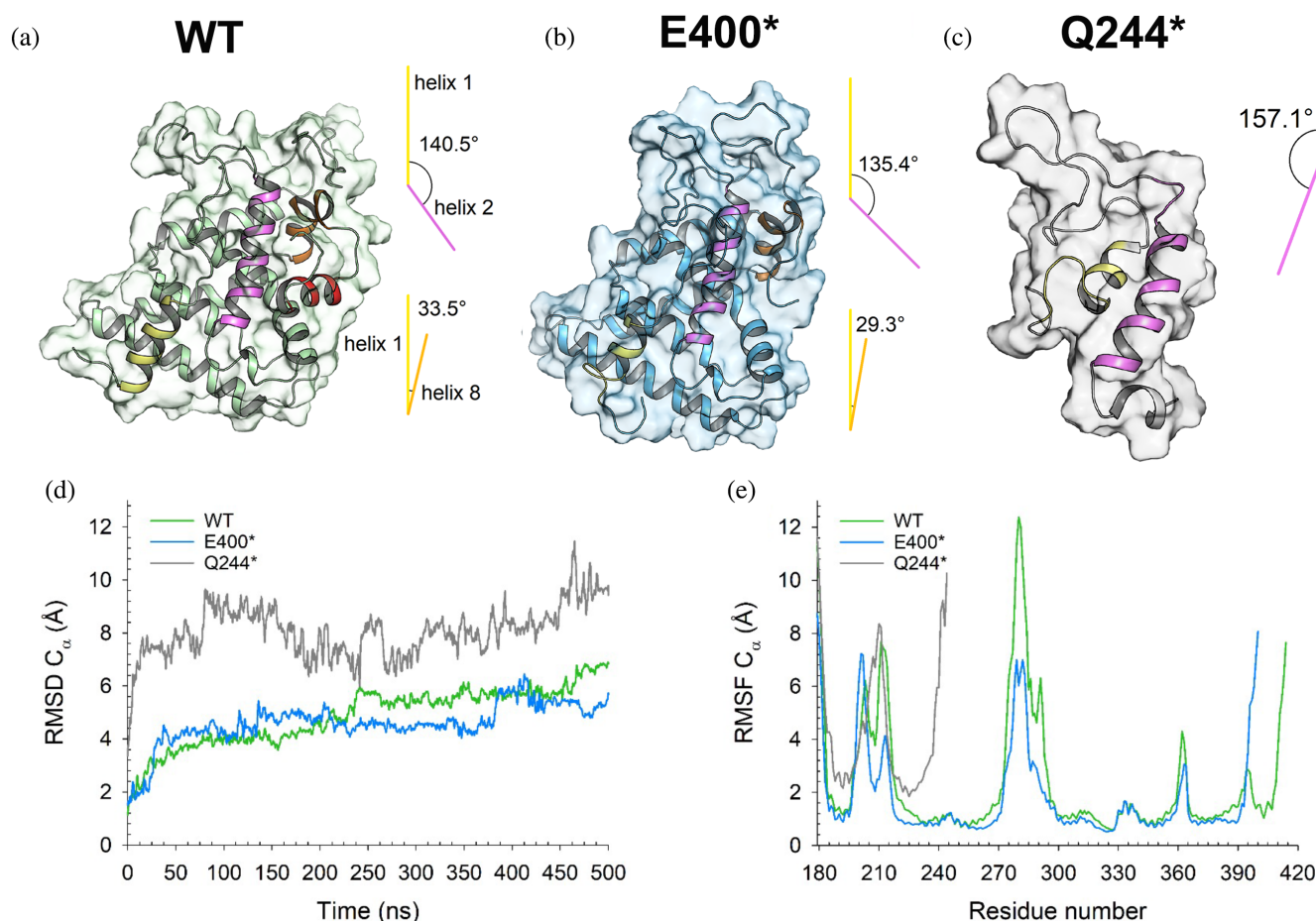
References	Variant (DNA)	Variant (protein)	DD	ID	Visual system defect(s) and visual deficit	EOE/S	ASD behavioral abnormalities	Hypotonia
RE20	c.954G>C	p.Glu318Asp	No but mild DMD	No (IQ 94; performance IQ 54)	OA; CVI; GVI	Atonic; Rolandic epilepsy	ASD	No
RE20	c.1117C>T	p.Arg373*	Yes, DMD	Mild (DQ ca. 60–70); speech delay	P/SOD; ONH; CVI; GVI	No	ASD	Yes
RE20	c.1217T>C	p.Met406Thr	Yes, DMD	Yes (IQ ND); speech delay	CVI; GVI	No	ASD	Yes
JS20	c.1217T>C	p.Met406Thr	Yes	Severe (IQ ND); speech delay; non-verbal	OA; suspected ON dysplasia; GVI	Seizures from 4mo	Short attention span	ND
JU21	c.599 C>G	p.Thr200Arg	Global	Yes (IQ ND); learning disability	Central, steady, maintained	No	ND	Yes
JU21	c.698G>A	p.Trp233*	Yes	Yes (IQ ND); speech difficulties	OA; microphthalmia; small ON head; CVI	No	Limited concentration and short attention span	ND
JU21	c.1024G>A	p.Glu342Lys	No	No	OA; ONH; LVA	No	ND	ND
JU21	c.1036_1047del	p.Glu346_Gln349del	Yes; walking delay	Yes (IQ ND); speech delay; dyslexia; learning disability	ONH; CVI; LVA	No	ND	ND
JU21	c.1115T>C	p.Leu372Pro	Yes; walking delay	Yes (IQ ND); speech delay	Small ON head; CVI; LVA	No	ND	ND
JU21	c.1115T>C	p.Leu372Pro	Yes; walking delay	Yes (IQ ND); speech delay	OA; ONH; CVI; LVA	One episode of FS	ND	ND
JU21	c.1115T>C	p.Leu372Pro	Yes; walking delay	Yes (IQ ND); speech delay	OA; CVI; LVA	No	ND	ND
JU21	c.1118_1123del	p.Arg373_Leu374del	No	ND	OA; ONH; LVA	No	ND	ND
JU21	c.1118_1123del	p.Arg373_Leu374del	No	ND	OA; ONH; LVA	No	ND	ND
JU21	c.1183G>A	p.Gly395Ser	Yes	Yes (IQ ND); learning disability	CVI; LVA	No	ND	Generalized
JU21	c.1198G>T	p.Glu400*	Yes	Yes (IQ ND); learning disability	OA; ONH; LVA	No	ASD; behavioral disorders	ND
BI21	c.854C>A	p.Ser285*	ND	ND	ND	ND	ND	ND

TABLE 1 (Continued)

References	Variant (DNA)	Variant (protein)	DD	ID	Visual system defect(s) and visual deficit	EOE/S	ASD behavioral abnormalities	Hypotonia
B121; STUDDER lab Unpublished	c.883T>C	p.Phe295Leu	No	No	OA; P/SOD; GVI	ND	ND	ND
B121	c.965T>A	p.Leu322His	ND	ND	ND	ND	ND	ND
B121	c.1168_1170del	p.Phe390del	ND	ND	ND	ND	ND	ND
STUDDER lab Unpublished	c.1184G>C	p.Gly395Ala	ND	Yes	GVI	Myoclonia	ASD	Yes
STUDDER lab Unpublished; FO19	c.1096C>T	p.Arg366Cys	Yes	Yes (speech delay)	ND	No	ND	ND

*Note:* Chronological order of publications describing BBSOAS cases and identified by their DNA and protein variant. Main clinical signs (columns) can include developmental delay (DD), intellectual disability (ID), visual system deficits such as optic atrophy (OA) and cortical/cerebral visual impairment (CVI), early-onset epilepsy and seizures (EOE/S), autism spectrum disorder (ASD) and behavioral abnormalities, and hypotonia. An extended version of these data, with additional columns showing the LOVD identifier of NR2F1 variants, clinical data about altered brain morphology as observed by MRI, and other less common clinical features, is available in Table S1. List of references (first column): BE20, (Bertacchi et al., 2020); BO14, (Bosch et al., 2014); CH16, (Chen et al., 2016); FO19, (Fokkema et al., 2019); JS20, (Jezela-Stanek et al., 2020); JU21, (Jurkute et al., 2021); KA17, (Kaiwar et al., 2017); SA13, (Sanders et al., 2012); RE20, (Rech et al., 2020); WA20, (Walsh et al., 2020); ZO20, (Zou et al., 2020). Abbreviations: ADHD, attention deficit hyperactivity disorder; ASD, autism spectrum disorder; DD, developmental delay; DMD, delayed motor development/poor coordination; DQ, developmental quotient; EOE/S, early-onset epilepsy/seizures; FS, febrile seizures; GVI, general visual impairments; ID, intellectual disability; IQ, intelligence quotient; IS, infantile spasms; LBD, ligand binding domain; LVA, low visual acuity; ND, not determined; OA, optic atrophy; ON, optic nerve; ONH, ON hypoplasia; P/SOD, pale/small optic disc; RB, repetitive behavior.





**FIGURE 2** Three-dimensional structure of NR2F1 LBD. (a) WT (green), (b) E400\* (blue), and (c) Q244\* (gray) variants after 500 ns MD simulations. Protein structure is shown as a cartoon with the molecular surface in transparency, helices 1 (residues 183–194), 2 (residues 219–236), 8 (382–394), and AF2 (residues 399–408) are colored in yellow, purple, orange, and red, respectively. Insets show the schematic representation of the angles between helix 1 and helix 2 (yellow and orange) and between helix 1 and helix 8 (yellow and red) and their relative values. (d) Root-mean square deviation (RMSD) of C $\alpha$  atoms calculated over 500 ns MD simulations with respect to the equilibrated structure of NR2F1 LBD WT (green), E400\* (blue), and Q244\* (gray). (e) Root-mean square fluctuation (RMSF) of C $\alpha$  atoms calculated over 500 ns MD simulations of NR2F1 LBD WT (green), E400\* (blue), and Q244\* (gray).

calculated on the same number of residues did not show any significant differences with that calculated on the full-length WT, as supported by the average values of  $4.87 \pm 1.12$  Å,  $4.97 \pm 1.18$  Å, and  $4.67 \pm 1.10$  Å displayed by full-length, Q244\*, and E400\* variants, respectively. Moreover, the Q244\* variant exhibited a major rearrangement of the N-terminal helices 1 and 2 with respect to the WT, as shown by the different relative orientation of the helices, whose angle significantly increased from  $140.5 \pm 7.7^\circ$  in the case of the WT to  $157.1 \pm 12.5^\circ$  for the Q244\* variant (Figure 2c). Concerning the E400\* variant, no hints of a potential unfolding were observed, as indicated by the preservation of the overall topology of the  $\alpha$ -helices. However, a small but significant reduction in the amplitude of the angle between helices 1 and 2 and between helices 1 and 8 was observed ( $140.5 \pm 7.7^\circ$  to  $135.4 \pm 7.7^\circ$  and  $33.5 \pm 6.4^\circ$  to  $29.3 \pm 7.7^\circ$ , respectively,

Figure 2b), suggesting that the truncation allosterically affects the N-terminal region, as well as the proximal helix 8.

We also evaluated the flexibility of the backbone of the LBD by monitoring the C $\alpha$  Root-Mean Square Fluctuation (RMSF). At odds with the Q244\* truncation, which showed the largest plasticity throughout the entire structure (Figure 2e), the E400\* variant displayed less flexibility at the level of loops 195–218 and 270–294 compared to the WT and negligible differences in the dimerization interface, again suggesting an allosteric effect due to the truncation that may not involve dimerization.

Taken together, our data suggest that both truncations are associated with BBSOAS through different molecular mechanisms. Since the Q244\* variant preserves only the DBD, the eventual DNA-binding process is rendered independent of LBD regulation; differently,

the E400\* variant acquires a more compact conformation, which is however unable to switch from the auto-repressed to the active form, leaving the CRS available to any potential binding partner without possible further regulation.

### 2.3 | LBD truncations and point mutations differently affect cell proliferation and survival

Having predicted the molecular effects of BBSOAS-associated LBD variants by structural analyses, we further evaluated the impact of selected NR2F1 mutations at the cellular level to gain insight into the NR2F1 structure–function relationship (Figure 3). We focused on the Q244\* and E400\* truncations, and five distinct point mutations selected with respect to (i) their spatial locations, that is, CRS (L252P), between CRS and DI (E318D), DI (G368D, L372P), and near AF2 (G395A) (Figure 3A), (ii) the diverse clinical symptoms with which they are associated (Tables 1 and S1), and (iii) their variable impact on isolated protein stability and affinity of both auto-repressed and active forms based on our computational analysis (Tables S2 and S3).

To investigate the effect of these pathogenic NR2F1 mutations on cell physiology, we took an *in cellula* over-expression approach by transfecting the different NR2F1 variants in HEK293T cells, which expressed marginal levels of endogenous NR2F1 (not shown). First, we quantified the fraction of NR2F1-positive cells entering different cell cycle phases by flow cytometry and double labeling with propidium iodide and NR2F1 (Figures 3b–d and S2a, b). Interestingly, HEK293T cells transfected with the Q244\*, E400\*, and L372P variants showed a significant decrease in S- and G2-phase entry when compared to WT NR2F1 (Figure 3b, d), indicating that distinct NR2F1 mutated forms can significantly inhibit cell cycle progression. Next, we assessed mitotic cells going through the M-phase with the help of the phospho-Histone-3 (PH3) antibody (Figure S2c). Upon transfecting the Q244\* and L372P forms, we found a significant decrease of PH3-positive cells in the NR2F1-positive population (Figure 3c), suggesting that cells carrying these variants fail to reach the M-phase and accomplish cell division. These findings reveal that distinct patient-specific LBD mutations affect cell cycle progression by slowing down the entry into the S-, G2-, and/or M-phase to different extents.

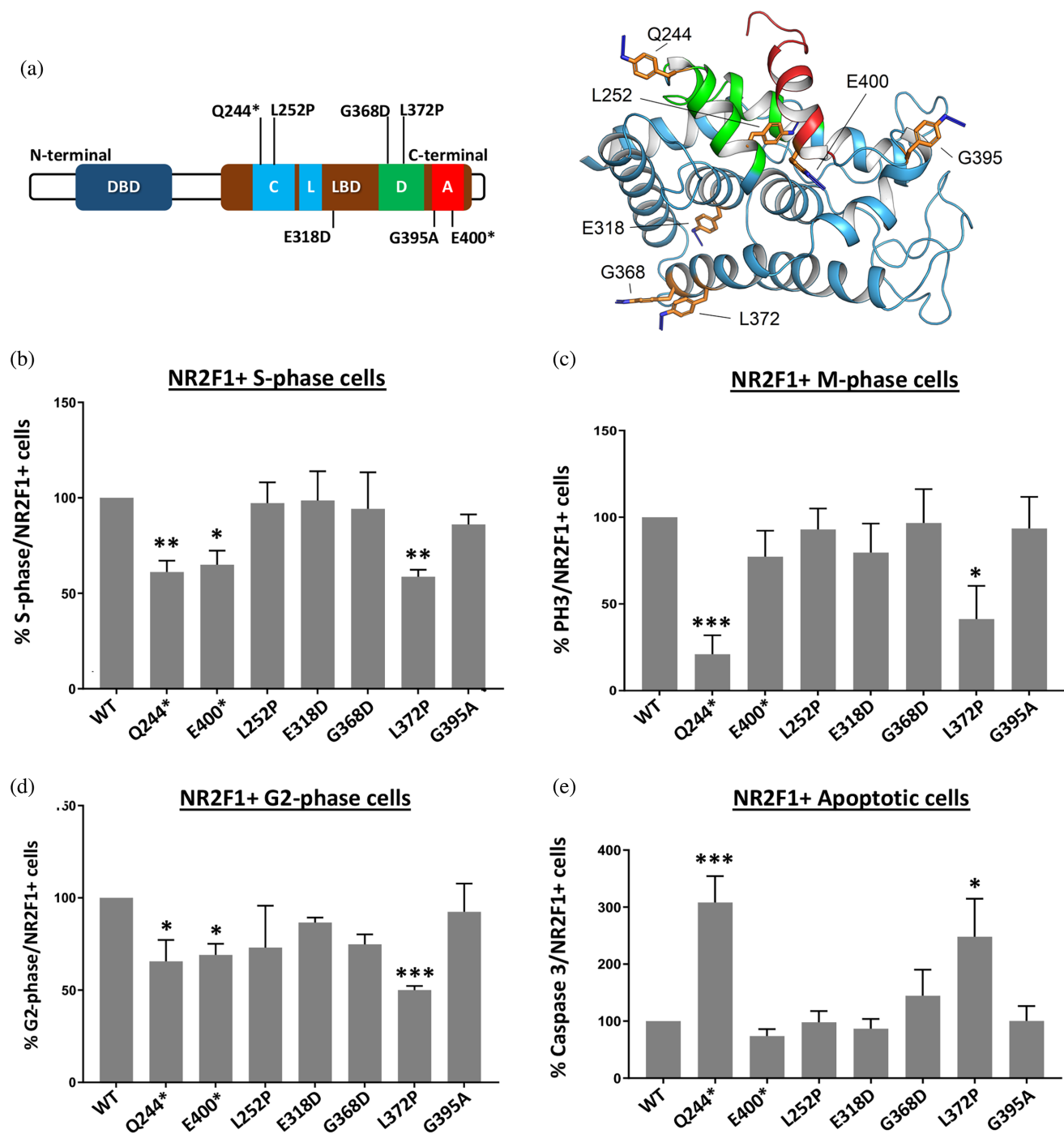
Next, we reasoned that a blockage in the cell cycle progression could be associated with a higher tendency to undergo cell apoptosis. Hence, we coupled the NR2F1 staining with cleaved Caspase-3 detection and quantified

the percentage of apoptotic cells in the NR2F1-positive fraction (Figure S2d). Interestingly, the same two variants that affected the M-phase entry (the Q244\* and the L372P) also significantly induced apoptosis (Figure 3e), displaying a 2.5- to 3-fold increase in apoptotic events compared to the control sample. Notably, the consequences of slowed cell cycle progression and increased apoptosis were readily visible 48 h after transfection, with the Q244\* and L372P variants inducing the most striking phenotype (Figure S3). Together, these results show that specific NR2F1 truncations and point mutations have direct consequences on cell proliferation and survival, that could partially explain how these NR2F1 variants impact human cell physiology. Of note, increased apoptosis could either result from impaired proliferation or from a direct control of cell-death pathway as part of a toxic *gain-of-function* mechanism.

### 2.4 | LBD truncations and point mutations can disturb the sub-cellular localization of NR2F1

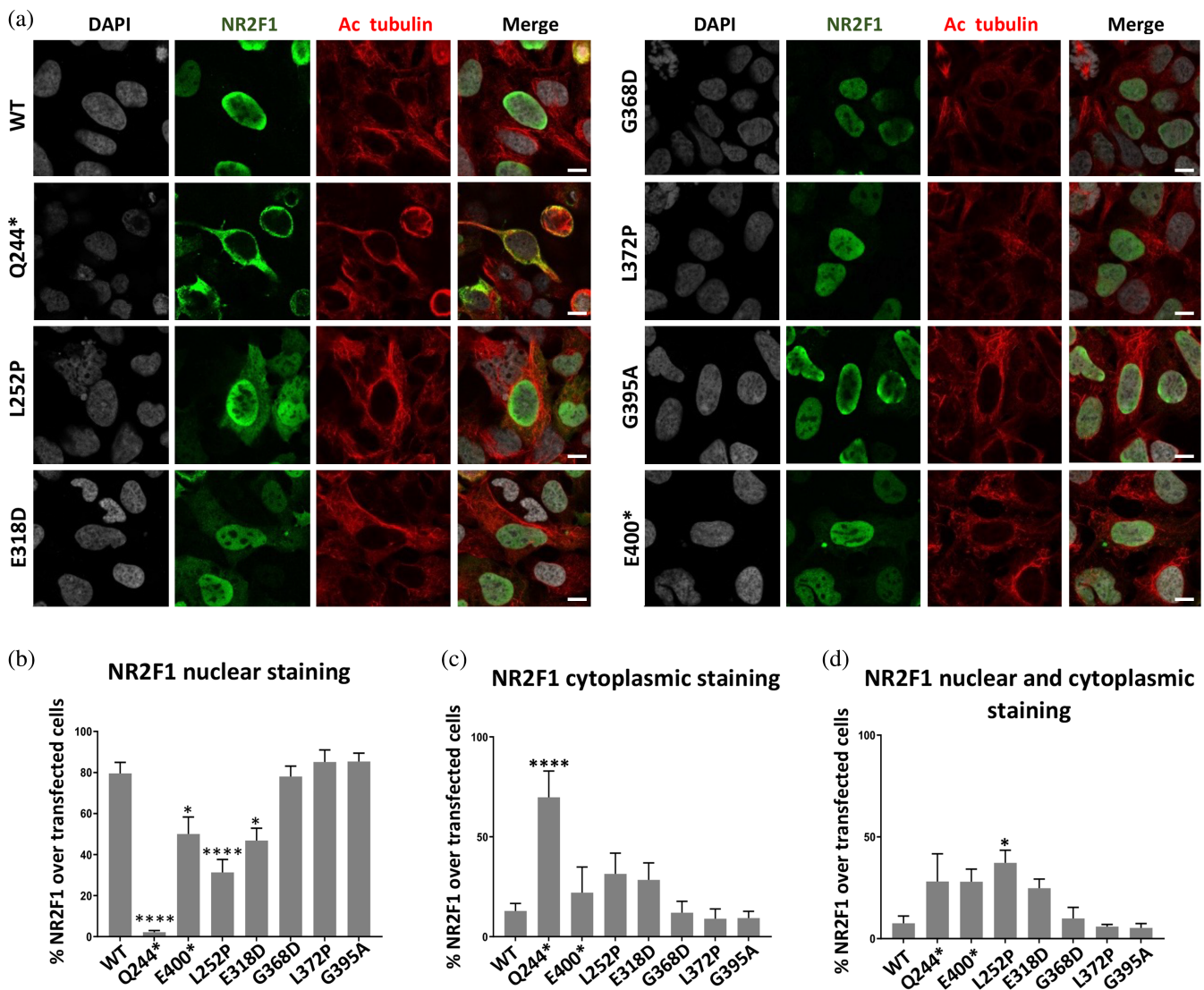
Based on the molecular and cellular changes due to distinct LBD mutations, we next assessed whether their presence could affect the normal sub-cellular localization of the protein in the nucleus. To this end, we co-immunostained NR2F1 in transfected HEK293T cells with DAPI (nuclear) and an anti-acetylated tubulin (cytoplasm/cytoskeleton) antibody and quantified the fraction of cells displaying NR2F1 either in the nucleus and/or cytoplasm (Figure 4). As expected from a transcriptional regulator, WT NR2F1 was mainly localized in the nucleus (Figure 4a), with approximately 80% of the cells showing this sub-cellular localization (Figure 4b). Strikingly, the vast majority (70%) of the Q244\*-transfected cells clearly displayed NR2F1 cytoplasmic staining (Figure 4a, c), while only a minor fraction (25%) revealed a partial NR2F1 localization both in the nucleus and the cytoplasm (Figure 4a, d). The L252P point mutation and, to a minor extent, the E318D and E400\* variants showed intermediate phenotypes with a partial accumulation of NR2F1 in both cytoplasm and nucleus (Figure 4a, d). On the contrary, other NR2F1 variants (G368D, L372P, and G395A) remained localized in the nucleus, similar to the NR2F1 WT (Figure 4a, b), suggesting that the presence of these pathogenic mutations does not affect sub-cellular distribution.

Taken together, these results show that only some LBD mutations/truncations affect NR2F1 nuclear localization. However, such abnormal intracellular distribution does not always mirror the presence of abnormal cell proliferation and survival. Our data suggest that proper



**FIGURE 3** The impact of NR2F1 LBD mutations on cell physiology. (a) Left: Schematic diagram of mutation positions investigated (C, putative coactivator recognition site, L, putative ligand-binding site, D, dimerization interface, A, AF2). Right: Three-dimensional structure of the LBD shown as a cyan cartoon with the AF2 helix shown in red and the CRS in green. Amino acid positions investigated are shown as orange sticks with N atoms in blue and labeled according to the original residue present in that position. (b–d) Cell cycle analysis upon transient transfection of HEK293T cells with the different NR2F1 variants as indicated. Histograms showing the percentages of S-phase (b), M-phase (c), and G2-phase (d) cells in NR2F1-positive populations. (e) Flow cytometry analysis of NR2F1-positive cells stained for anti-cleaved-caspase 3 and undergoing apoptosis. FACS gatings are presented in Figure S3. Means  $\pm$  SD of at least three independent experiments (each including 10,000 gated cells per sample) are shown (\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.0005$ , one-way ANOVA with Dunnett's test post-hoc analysis). All data were normalized on the control condition (WT; 100%) and percentages are calculated over the NR2F1+ (NR2F1-transfected) cell fraction.





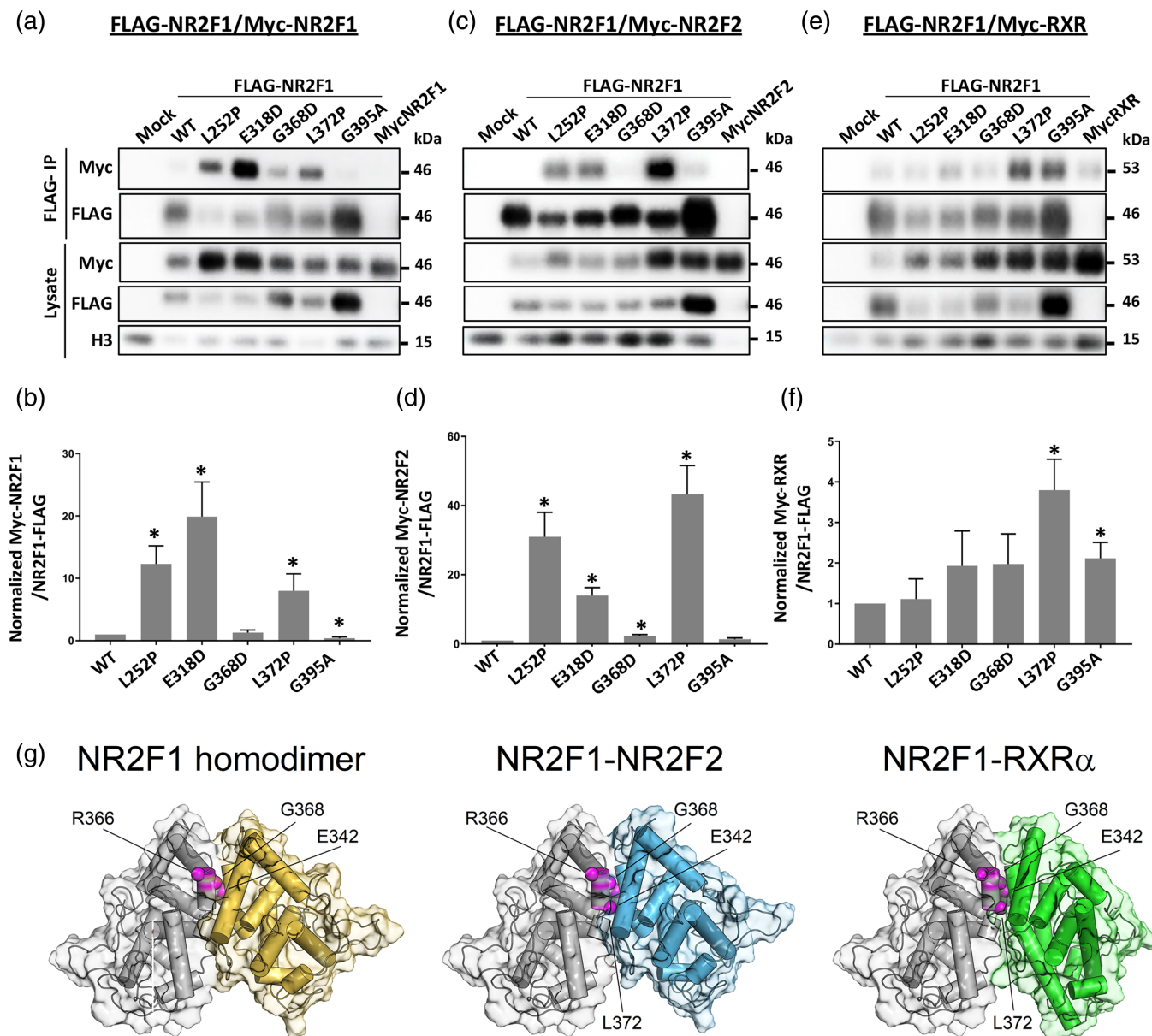
**FIGURE 4** The impact of NR2F1 LBD mutations on NR2F1 sub-cellular localization. (a) Confocal images showing NR2F1 (green) and acetylated tubulin (red) immunofluorescence staining on HEK293T cells transiently overexpressing NR2F1 (WT or mutants), as indicated on the side. (b–d) Histograms showing the quantification of the percentage of NR2F1 sub-cellular location in the nucleus (b), in the cytoplasm (c), and in both the nucleus and cytoplasm (d). Only NR2F1-positive cells were counted (manual quantification in ImageJ). Means  $\pm$  SD of three independent experiments are shown ( $*p < 0.05$ ,  $****p < 0.0001$ , two-way ANOVA with Dunnett's test post hoc analysis), including a total number of at least 700 cells per condition/staining. Scale bars: 10  $\mu$ m.

nuclear localization alone is not indicative of proper NR2F1 functions, since pathogenic mutations can exert toxic effects when accumulated in the cytoplasm (as for Q244\*), as well as when normally localized in the nucleus (as for L372P).

## 2.5 | Impact of LBD mutations on the formation of NR2F1-protein complexes

Since nuclear receptors normally undergo homo- or heterodimerization with other nuclear receptors before or upon DNA binding *via* the LBD domain (Chandra

et al., 2017; Lou et al., 2014) (modeled in Figure 1d), we next assessed the impact of NR2F1 pathogenic variants on the capacity of forming dimers. In particular, we evaluated the effect of LBD mutations on the ability of NR2F1 to dimerize with itself, with NR2F2 and/or with RXR $\alpha$  (Evans & Mangelsdorf, 2014). To first investigate NR2F1 homodimerization, we co-expressed FLAG-tagged NR2F1 (FLAG-NR2F1-WT or mutants) and Myc-tagged WT NR2F1 (Myc-NR2F1) in HEK293T cells followed by co-immunoprecipitation (co-IP) with anti-FLAG antibody and immunoblotting. The detection of Myc-NR2F1 co-IP with FLAG-NR2F1 on the immunoblot would imply a complex formation (e.g., homodimers or oligomers).



**FIGURE 5** The impact of NR2F1 LBD mutations on its oligomerization in cells. (a, c, e) Immunoblots of lysates from transiently co-transfected HEK293T cells with NR2F1-FLAG (WT or mutants) and Myc-NR2F1 (a), Myc-NR2F2 (c), or Myc-RXR $\alpha$  (e). (b, d, f) Densitometric analysis of co-IP samples from experiments as shown in (a), (c), and (e) using ImageJ software. Normalized co-IP Myc-tagged partner intensity is shown (see Section 5). Means  $\pm$  SD of at least three independent experiments are shown (\* $p < 0.05$ , unpaired  $t$ -test, compared to NR2F1-WT). (g) Results from docking simulations of NR2F1 LBD with itself (left), NR2F2 LBD (center) and RXR $\alpha$  LBD (right), proteins are represented as cartoon with cylindrical helices together with the respective molecular surface in transparency. NR2F1 LBD is shown in gray and yellow, NR2F2 LBD in blue and RXR $\alpha$  LBD in green; the C $\alpha$  of the residues belonging to the dimer interface and associated with BBSOAS are represented by magenta spheres and labeled.

Surprisingly, we found that the co-IP of Myc-NR2F1 with FLAG-NR2F1-WT was barely detectable (Figure 5a), as with the FLAG-NR2F1 G368D and G395A variants. On the contrary and surprisingly, significant amounts of Myc-NR2F1 co-IP with the FLAG-NR2F1-L252P, E318D, and L372P variants were observed, even though expression levels of these variants were consistently lower than that of the WT. By performing densitometric analysis

followed by normalization of the Myc-NR2F1 by FLAG-NR2F1 co-IP levels, we found that L252P, E318D, and L372P variants increased the complex formation by 12-, 20-, and 8-fold, respectively, when compared to WT NR2F1 (Figure 5b). Notably, the expression of the G395A variant consistently resulted in a reduced NR2F1/NR2F1 complex formation although its expression level was always higher than that of the WT NR2F1.



Next, we performed similar experiments to investigate the NR2F1 heterodimerization with its homolog NR2F2 and with the nuclear receptor RXR $\alpha$ , previously described as potential NR2F1 dimerization partners (Cooney et al., 1992; Pinaire et al., 2000; Zhang & Dufau, 2001), by co-expressing FLAG-NR2F1 (WT or mutants) with Myc-tagged NR2F2 (Myc-NR2F2) or Myc-tagged RXR $\alpha$  (Myc-RXR $\alpha$ ) (Figure 5c–f). As with the NR2F1 homodimerization data, we could not readily detect a WT NR2F1/2 interaction (Figure 5c). However, we observed a 31-, 14-, and 43-fold increase in the NR2F1/2 complex upon the expression of L252P, E318D, and L372P variants, respectively (Figure 5d). A small but significant increase ( $\sim$ 2-fold) in NR2F1/2 co-IP was also detected in the presence of the NR2F1-G368D, but not the G395A variant. Consistently, while we revealed a  $\sim$ 3-fold increase in RXR $\alpha$  interaction with NR2F1-L372P, no significant RXR $\alpha$  interaction with either WT NR2F1 or NR2F1-G368D above the background level was detected (Figure 5e, f). However, unlike NR2F1/NR2F1 or NR2F1/2 interactions, the NR2F1 variants L252P and E318D did not significantly enhance NR2F1/RXR $\alpha$  interaction. Finally, while the expression of the G395A variant reduced the NR2F1 homo-oligomerization by about  $\sim$ 60% (Figure 5b), it significantly increased the NR2F1/RXR $\alpha$  interaction by approximately 2-fold (Figure 5f).

Overall, our data in HEK293T cells suggest that NR2F1 homo- and heterodimerization may not always be a prevalent event and that some NR2F1 pathogenic variants may instead promote the protein propensity to form homo-/heterodimers and/or higher order NR2F1 complexes. Furthermore, depending on the identity of the dimerization/oligomerization partners, different variants may exert varying degrees of impact on their interactions.

## 2.6 | *In silico* NR2F1 protein complexes modeling

To further interpret the above-described experimental results, we predicted the quaternary structure of NR2F1 variants *in silico*. Since no experimental structure of NR2F1 dimers was available, we used PIPER (Kozakov et al., 2006) to run unbiased protein–protein rigid-body docking simulations. Simulations revealed a dimeric assembly in line with the conserved DI of other proteins belonging to the same nuclear hormone receptor superfamily, namely NR2F2, RXR $\alpha$  (subfamily NR2), RAR $\beta$ , and LXR $\beta$  (subfamily NR1). The similarity concerned both the sequence (Figure S1) and the physicochemical nature of the interactions between monomers (Figure S4). Based on the structural information provided by the NR2F2 homodimer (Kruse et al., 2008), RXR $\alpha$

homodimer (Egea et al., 2002), RXR $\alpha$ -RAR $\beta$  (Chandra et al., 2017), and RXR $\alpha$ -LXR $\beta$  (Lou et al., 2014) heterodimers (Figure S4), we predicted a reliable three-dimensional structure of the NR2F1 homodimer and heterodimeric complexes of NR2F1 with NR2F2 and RXR $\alpha$  (Figure 5g), both in the auto-repressed and active forms. To probe the robustness of predicted NR2F1 LBD homodimers, rigid-body docking simulations were performed using ZDOCK 2.3 (Chen et al., 2003), whose score (ZDOCK-score) was demonstrated to correlate with the experimental binding affinity upon an exhaustive sampling of the roto-translational space and under the assumption of rigid body-like binding (Dell’Orco et al., 2007). We identified 23 poses for the LBD, in both states, resembling the docked complex obtained with PIPER (native-like poses, C $\alpha$ -RMSD  $<$ 1 Å, Table S4). The active form presented higher-ranked best solutions compared to the auto-repressed form in all four docking runs (Table S4), which reflected the higher average ZDOCK-score of the native-like solutions of the active form ( $58.06 \pm 6.63$  vs.  $49.31 \pm 3.97$ , Table S4). Although no direct estimate of the binding free-energy was possible, since the size of the protein–protein interface of NR2F1 homodimers significantly exceeded that of the complexes used in the original affinity-ZDOCK-score correlation (Dell’Orco et al., 2007), this result suggests that the auto-repressed form would dimerize with significantly lower affinity compared to the active one. In summary, independent and unbiased docking simulations suggest, for the NR2F1 LBD, a dimerization process substantially in line with that of other nuclear receptors in solution, both structurally and in terms of affinity (Figure 5g).

Based on the predicted structural organization of NR2F1 homo- and heterodimers, we then calculated the effects of BBSOAS-variants relative to the WT on quaternary complex structures, in terms of folding stability ( $\Delta\Delta G_f^{\text{app}}$ ) and binding affinity ( $\Delta\Delta G_b^{\text{app}}$ ). Among all tested pathogenic variants, except for T200R, the free energy of folding  $\Delta\Delta G_f^{\text{app}}$  of the homo/heterodimers, with respect to the WT in both the auto-repressed and active conformation, was increased, regardless of the dimerization partner (see Video S3 for results calculated on NR2F1 homodimer in heterozygosis as an example). This suggests a destabilization of the LBD dimeric complexes in the presence of BBSOAS-associated variants, with the largest effect displayed by the NR2F1 homodimer with mutations in a potential homozygote genetic condition (Tables S2 and S3).

Moreover, we predicted a lower effect on the dimerization affinity for BBSOAS-associated variants, as only R366C, G368D, and L372P displayed significantly diminished affinity (i.e., higher  $\Delta\Delta G_b^{\text{app}}$  values) for each of the tested complexes, in both the active and the auto-

repressed conformations, although with numerical differences among partner receptors. This is not surprising, since all three residues belong to helix 7, which forms the DI together with helices 6 and 8. Furthermore, such mutations displayed an overall more prominent decrease in affinity in the case of the active form (Table S3) compared to the auto-repressed one (Table S2), most likely due to the increased interaction surface. Notably, the E342K mutation turned out to be detrimental only for the active form, where the larger DI brings residue 342 in close contact with the positively charged residues K308, K369, and R373 of the other subunit. The E342K variant would therefore replace a negative charge with a positive one, thus accounting for electrostatic repulsion to cause decreased dimer affinity. All other tested variants showed a negligible effect on dimer affinity, with  $\Delta\Delta G_b^{\text{app}}$  values never exceeding  $\pm 1$  kcal/mol (Tables S2 and S3).

Taken together, our results suggest that the predicted destabilizing effects of the point mutations on the isolated forms of NR2F1 correlate well with the propensity of NR2F1 to form dimers and/or oligomers in cells; however, dimeric interfaces that differ from the canonical one experimentally observed in crystal structures of other nuclear receptors might nevertheless form under specific conditions.

## 2.7 | Participation of the BBSOAS mutation positions in the dimerization revealed by GCE-enabled photocrosslinking in living cells

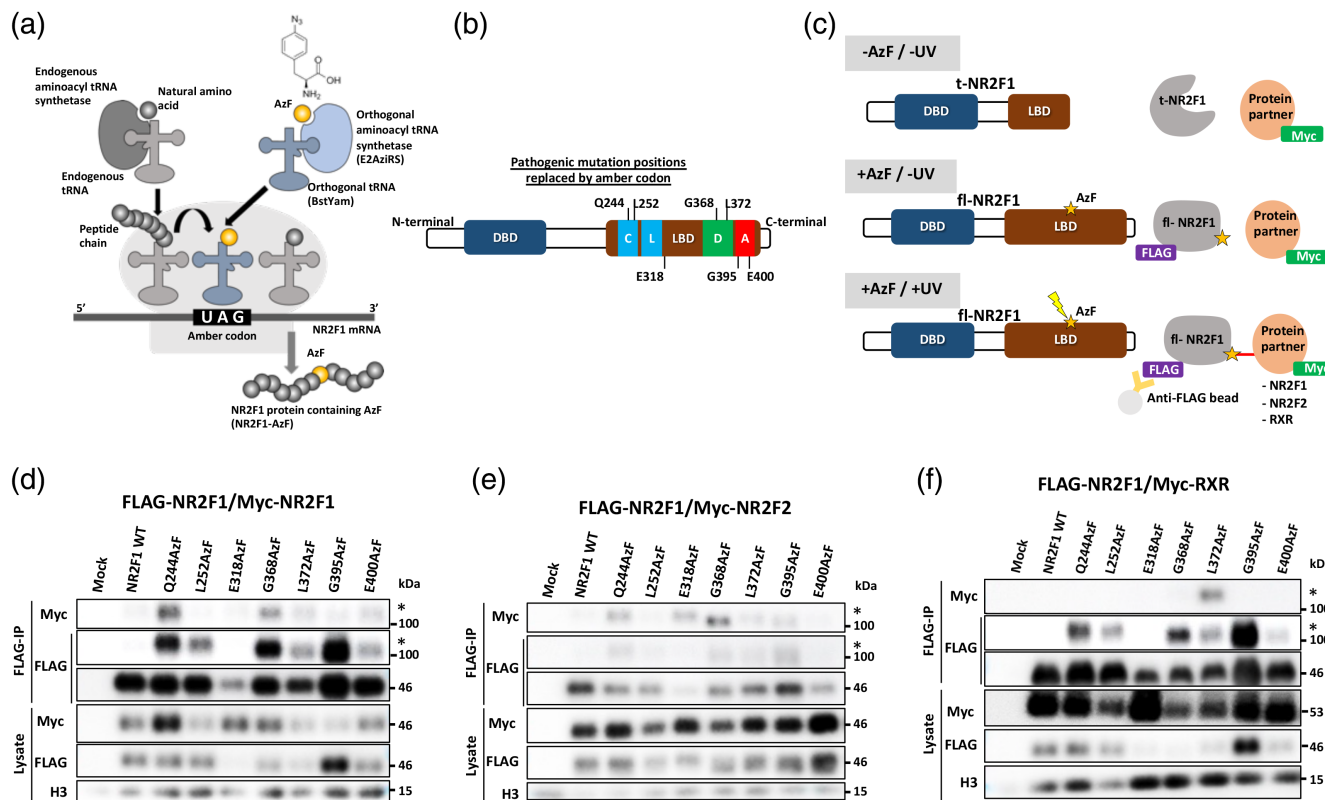
Several putative protein interaction sites in the LBD should allow NR2F1 to dynamically work in concert with various protein partners depending on the cellular stages and context. However, our conventional co-IP approach could not easily detect homo- and/or heterodimerization of predicted partners with the WT NR2F1 form despite the *in silico* prediction of the event (Figure 5), indicating the limitations of the co-IP approach to stably capture transient interactions in living cells.

To overcome these experimental constraints and more precisely study protein interactions at selected BBSOAS mutation positions, we applied the genetic code expansion (GCE) technology to co-translationally and site-specifically incorporate a small photo-crosslinking amino acid, *p*-azidophenylalanine (AzF) into the LBD of NR2F1 (Figure 6a). This was achieved by co-transfecting HEK293T cells with NR2F1 isoforms (*NR2F1am*), each carrying an amber codon (*TAG*) at a desired position, along with an amber suppressor tRNA derived from *B. stearothermophilus* tRNA<sup>Tyr</sup> (Bst-Yam) and an enhanced aminoacyl tRNA synthetase (E2AziRS) in the

presence of AzF (Seidel et al., 2017). A subsequent UV irradiation allowed a photo-activated covalent crosslinking between AzF in the LBD and its interactors within the radius of reach of AzF ( $\sim 9$  Å from C $\alpha$ ) within a living cell (Seidel et al., 2017). We hypothesized that variable pathogenic effects of the BBSOAS mutations in different LBD protein-binding sites might be due to their interferences with different protein–protein interactions. Therefore, we focused our investigations on the same mutation locations at the protein interaction sites where we investigated the cellular effects (Figures 3 and 4).

To this end, we generated a series of C-terminal FLAG-tagged full-length NR2F1 variants where the amino acid positions of variants Q244\*, L252P, G318D, G368D, L372P, G395A, and E400\* were replaced by AzF in response to the amber codon (*TAG*) (Figure 6a, b). The incorporation of AzF was successful with minimal read-through in the absence of AzF, and NR2F1-interacting proteins could be covalently captured by photocrosslinking *via* AzF (Figure S5). Then, we compared the ability of different AzF-NR2F1 variants to capture the elusive dimerization *in cellula* by co-expressing FLAG-AzF-NR2F1 with Myc-NR2F1, Myc-NR2F2, or Myc-RXR $\alpha$  in HEK293T cells. GCE-enabled photocrosslinking followed by FLAG-NR2F1 co-IP was subsequently performed (strategy in Figure 6c). The data showed that the NR2F1 homodimerization and NR2F1/2 heterodimerization were most efficiently captured by Q244AzF and G368AzF variants, while the NR2F1/RXR $\alpha$  heterodimerization was mainly captured by the L372AzF one (Figure 6d–f). Although some NR2F1/RXR $\alpha$  heterodimerization could be captured by the Q244AzF variant, this was less efficient (not shown). To further probe the conformational impact of pathogenic variants on the dimer formation while avoiding possible functional interference by the photocrosslinker incorporation, we chose to place the photocrosslinker at the residue 244 (Q244AzF), since this residue was not predicted as a critical residue for dimerization (Figures 1d and 5g), but close enough to allow the capture of dimerization partners of WT NR2F1 without perturbing the cellular function of the protein (Figure S6).

To investigate the effects of point mutations L252P, E318D, G368D, L372P, and G395A on dimer formation using the photocrosslinking *via* the Q244AzF, we co-expressed full-length FLAG-NR2F1 or FLAG-Q244AzF variants (with or without the L252P, E318D, G368D, L372P, or G395A mutations) with Myc-tagged NR2F1, NR2F2, or RXR $\alpha$ , and the *E2AziRS*/Bst-Yam pair in the presence of AzF (Figure 7). After photocrosslinking with UV, we performed co-IP of the cell lysates with the anti-FLAG antibody and immunoblotting. As expected, the photocrosslinking *via* the Q244AzF successfully captured



**FIGURE 6** Covalent and site-specific NR2F1 dimeric partner capturing *via* GCE. (a) Principles of site-specific incorporation of photocrosslinking amino acid by amber codon suppression. Endogenous aminoacyl tRNA synthetase charges the endogenous tRNA with its cognate amino acid. Aminoacyl-tRNA enters the ribosome and adds the amino acid to the corresponding codon. For the AzF incorporation in NR2F1, the orthogonal aminoacyl tRNA synthetase (E2AzIRs) catalyzes the aminoacylation between AzF and the orthogonal tRNA (BstYam). The AzF-charged tRNA enters the ribosome and incorporates the AzF in response to the designated amber codon. The translation continues and produces full-length, AzF-containing NR2F1 protein. (b) Schematic diagram of pathogenic mutation positions replaced by amber codon for AzF incorporation. (c) Diagram showing UV-induced, site-specific crosslinking *via* AzF. NR2F1 is produced as a truncated form (t-NR2F1) in the absence of AzF due to the designated amber codon. In the presence of AzF, the amber codon suppression allows the incorporation of AzF at the selected amber codon position and the NR2F1 protein carrying AzF is translated into the full-length FLAG-tagged form (fl-NR2F1). Upon UV irradiation (365 nm), AzF forms a covalent bond between FLAG-NR2F1 and Myc-tagged putative partners (NR2F1, NR2F2, RXR $\alpha$ ). The proteins in the crosslinked complex can be co-IP with NR2F1 by using anti-FLAG antibody-conjugated beads and detected in a higher molecular weight complex by immunoblot. (d–f) Immunoblots showing site-specific capturing of homodimer (FLAG-NR2F1/Myc-NR2F1 (d)), heterodimers (FLAG-NR2F1/Myc-NR2F2 (e)), and FLAG-NR2F1/Myc-RXR $\alpha$  (f). The expected position of the covalently bound dimers is denoted by an asterisk. Data shown are representative of three independent experiments.

the elusive NR2F1 homodimer and NR2F2 and RXR $\alpha$  heterodimers, as indicated by an enrichment of the FLAG-Q244AzF-NR2F1/Myc-NR2F1, NR2F2, or RXR $\alpha$  complex at the molecular weight corresponding to the NR2F1 homodimer or heterodimer (Figure 7a–c). Notably, the presence of L252P, E318D, G368D, and L372P mutations clearly led to a reduction in NR2F1 protein expression, although it was still detectable after immunoprecipitation. Densitometric analysis followed by normalizing Myc-NR2F2 by FLAG-NR2F1 co-IP indicated that the G368D mutation significantly reduced the ability of NR2F1 to form the homodimer (Figure 7a) as well as the heterodimer with NR2F2 and RXR $\alpha$  (Figure 7b, c). This result confirmed the *in silico* prediction that this

mutation would exert the strongest detrimental effect on dimer stability and affinity (Tables S2 and S3). Moreover, a statistically significant effect of the L252P, E318D, and L372P variants on the homo- and heterodimer formation was not detected even though the conventional NR2F1 co-IP showed that these mutations favored complex formation (Figure 5a–c). These observations suggest that the structural destabilizing effect of these pathogenic variants not only led to the reduction of the NR2F1 protein expression but also increased the propensity of NR2F1 to form unexpected homo- and hetero-dimeric/oligomeric complexes. Interestingly, of all the mutations tested, the G395A was the only one that caused a strong increase in NR2F1 protein expression level. Although this mutation

did not significantly affect the NR2F1 homodimerization or heterodimerization with RXR $\alpha$  (Figure 7a, c), it significantly reduced the NR2F1/2 heterodimerization

(Figure 7b). Taken together, the covalent and site-specific protein-protein interaction captured by GCE-enabled photocrosslinking allowed us to discern the effects of

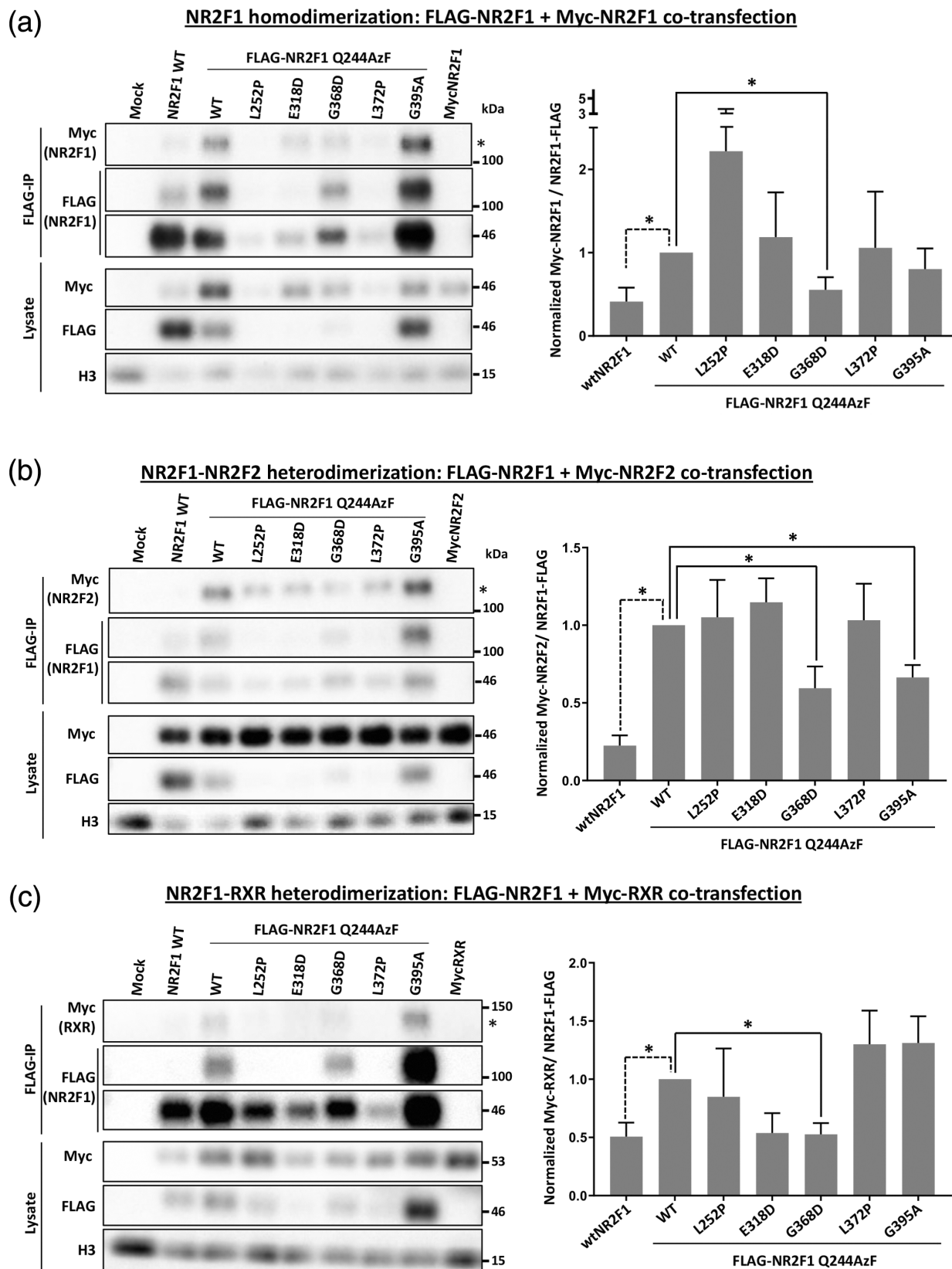


FIGURE 7 Legend on next page.



different pathogenic variants on distinct NR2F1 dimeric pairs in living cells.

## 2.8 | CRABP2 as a novel partner of NR2F1

The abnormal cellular localization of some NR2F1 variants led us to hypothesize that the nuclear transport of NR2F1 might require co-factor binding, which could be disrupted by the presence of specific LBD mutations (Figure 4). In light of its sequence similarity with RXR $\alpha$  and RAR $\beta$  (Figure S1) and of the detected NR2F1 heterodimerization with RXR $\alpha$  in living human cells (Figure 7), we hypothesized that NR2F1 might be directly or indirectly involved in retinoic acid (RA)-induced regulation, as previously suggested (reviewed in). The cellular retinoic acid-binding protein 2 (CRABP2), which is responsible for the nuclear transport of RA to nuclear receptors, such as RXR and RAR (Sessler & Noy, 2005), represented a potential NR2F1 partner in this process. However, the dynamic shuttling of CRABP2 between the cytoplasm and the nucleus constituted a challenge in detecting transient interactions in living cells. By taking advantage of the GCE-enabled photocrosslinking *via* the Q244AzF-NR2F1 followed by endogenous CRABP2 co-IP and immunoblotting, we explored the potential interaction between NR2F1 and CRABP2 (Figure 8). We found that CRABP2 could be covalently captured *via* photocrosslinking by Q244AzF-NR2F1, revealing the *in cellula* direct interaction between the two molecules through the CRS of NR2F1 (Figure 8a). Using the same FLAG-tagged NR2F1 mutations described above, we further examined the impact of pathogenic variants on CRABP2/NR2F1 interaction. Notably, the overexpression of the G368D and G395A mutations led to a marked increase in the NR2F1/CRABP2 complex, while an opposite trend was observed for the L252P and L372P variants. The expression of the E318D mutation, however, did not affect the extent of NR2F1/CRABP2 complex formation

(Figure 8b). To support the physiological significance of this interaction, we performed an immunofluorescence of NR2F1 and CRABP2 in the embryonic (E)13.5 mouse forebrain and developing eye, two anatomical sites known to be affected in *Nr2f1* mutant mice and perturbed in BBSOAS patients (Bertacchi et al., 2022; Tocco et al., 2021). Strikingly, a clear NR2F1 and CRABP2 protein co-expression was observed in cells of the medial-posterior cortex from which the hippocampus will form, and of the developing neural retina (Figure 8c, d), supporting an *in vivo* physiological interaction between NR2F1 and CRABP2 during mouse development.

Finally, to identify potential molecular interaction interfaces between CRABP2 and NR2F1, unbiased molecular rigid-body docking simulations were performed. The highest-scored docking poses of apo and holo CRABP2 (Vaezeslami et al., 2006) with NR2F1 displayed substantially the same binding interface, suggesting that the interaction between NR2F1 and CRABP2 may occur regardless of the loading state of the RA transporter and without major conformational changes in either protein (Figure 8e). Moreover, the hydrophobic pockets surrounding RA in NR2F1 and CRABP2 were found to be aligned in the docked complex (Figure 8e), thus allowing RA to be directly channeled from CRABP2 to NR2F1 similarly to what occurs in the RA transport to RAR (Budhu & Noy, 2002). Interestingly, *in silico* predictions of the effects of the experimentally tested point mutations on the stability of the docked NR2F1/CRABP2 assembly resulted in line with those obtained by GCE-enabled photocrosslinking: a high destabilization predicted for the L252P and L372P variants ( $\Delta\Delta G_f^{\text{app}} = 57.99$  and  $57.00$  kcal/mol, respectively), a stabilizing effect for the G368D variant ( $\Delta\Delta G_f^{\text{app}} = -4.23$  kcal/mol) and a negligible effect for the G395A variant ( $\Delta\Delta G_f^{\text{app}} = 1.55$  kcal/mol). Taken together, these molecular data support the physiological and pathological significance of the newly discovered interaction between the CRS of NR2F1 and CRABP2 in human cells and mouse embryos.

**FIGURE 7** The impact of NR2F1 LBD mutations on its dimerization in living cells. (a–c) Panels showing immunoblots of the immunoprecipitated samples and whole lysates with anti-FLAG antibody. The pIRE4-Azi plasmid containing E2AziRS and BstYam, pcDNA3.1 NR2F1Q244\*-FLAG without or with pathogenic mutations, and the plasmid carrying the gene of each putative protein partner (Myc-NR2F1, Myc-NR2F2, or Myc-RXR $\alpha$ ) were transiently cotransfected in HEK293T cells. Left: Immunoblots show the dimerization between FLAG-NR2F1 and its putative partners Myc-NR2F1 (a), Myc-NR2F2 (b), and Myc-RXR $\alpha$  (c). The asterisk indicates the approximated position of the dimer's molecular weight. Right: Quantitative analysis of the impact of the pathogenic mutations on the dimeric binding between NR2F1 and NR2F1 (a), NR2F1 and NR2F2 (b), and NR2F1 and RXR $\alpha$  (c). The densitometric analysis and normalization were performed as described in Figure 5. Means  $\pm$  SD of three independent experiments are shown (\* $p < 0.05$ , unpaired *t*-test, compared to NR2F1-Q244AzF).



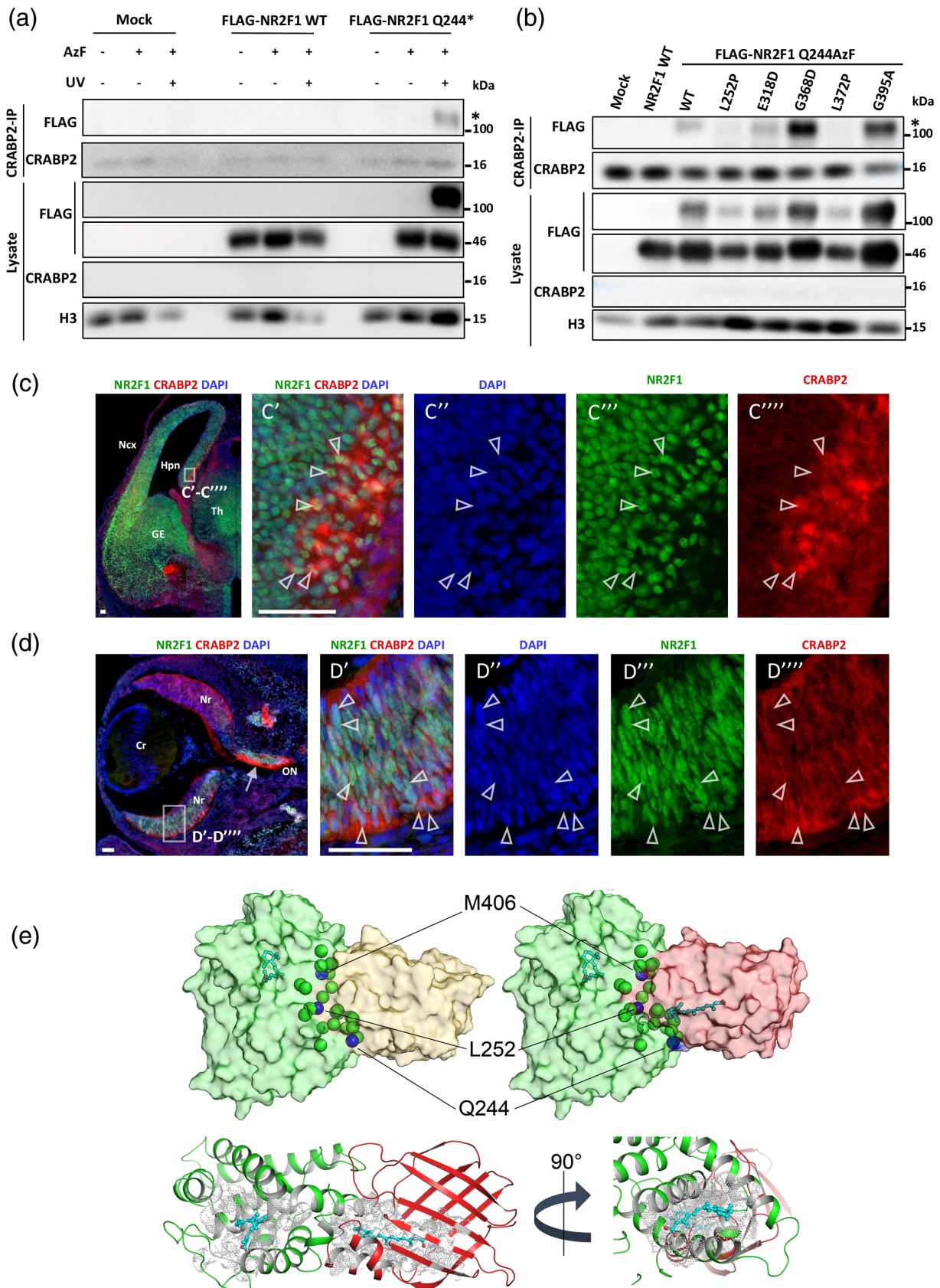


FIGURE 8 Legend on next page.

### 3 | DISCUSSION

The functions of NR2F1 as a transcriptional activator or repressor is achieved through complex protein–protein interactions involving coactivators, corepressors, and other transcription factors (Bertacchi et al., 2019; Montemayor et al., 2010; Pereira et al., 2000). In this study, we unveiled the functional consequences of LBD BBSOAS pathogenic variants by assessing NR2F1 basic cellular functions and integrating them with the structural analysis and GCE-assisted photocrosslinking in living cells. Thanks to this interdisciplinary work, we could unveil (i) the diversity of quaternary conformations of full-length NR2F1 in contact with different protein partners thanks to *in silico* and *in cellula* approaches, (ii) the variable conformations of NR2F1 variants, (iii) the disrupted interplay with dimeric partners and (iv) the newly identified co-factor, CRABP2.

#### 3.1 | Dimerization of NR2F1 as a dynamic and transient process in living cells

Depending on partners and the cellular context, nuclear receptor dimers can trigger different regulatory events by binding to distinct sequences of target genes and controlling various processes (Amoutzias et al., 2008). However, little is known about the role of LBD in NR2F1 dimerization due to the lack of structural information. Our thorough *in silico* modeling of NR2F1 dimers based on the dimeric structures of other nuclear receptors has allowed the prediction of the impact of BBSOAS mutations on

LBD dimerization, some of which were validated in human cells. Noteworthy is that the predicted changes in affinity and stability of the NR2F1 dimer attributed to the pathogenic substitutions are highly correlated ( $R^2 = 0.807$  for NR2F1 homodimer in heterozygosis). The significant correlation between mutation-induced variations in dimerization affinity and folding stability in the LBD suggests that the pathogenic mechanisms underlying BBSOAS may also involve a synergy between the LBD dimerization process (pre-/post-DNA binding) and its functional properties.

Early studies of NR2F1 in solution set the widely referenced notion that the dimerization of NR2F1 is a prerequisite for its DNA-binding ability and function (Sagami et al., 1986, Cooney et al., 1992). However, the necessity of the pre-formed NR2F1 dimer in the cell is still debated, since it has been observed in solution that NR2F1 monomer could also bind to response elements then dimerize, albeit less efficiently (Cooney et al., 1992). Indeed, direct evidence of a stable dimer formation prior to DNA binding is lacking. Data mining, manual curation, and a data integration study reported 15 years after these early studies did not find further reports of direct evidence of functional NR2F1 homodimers (Amoutzias et al., 2007). Thus, physical evidence that the stable pre-formed NR2F1 homodimer is required for its function in living cells had not been clearly demonstrated.

Our data in human cells show that the NR2F1 WT dimers were not readily detectable by NR2F1 traditional co-IP. Remarkably, pathogenic mutations that were predicted to strongly destabilize monomeric NR2F1 enhanced the NR2F1-NR2F1 and NR2F1/2 dimerization and/or oligomerization in cells. This strongly suggests

**FIGURE 8** Direct interaction between NR2F1 and CRABP2. (a) HEK293T cells were transiently co-transfected with plasmids containing E2AziRS, BstYam, and FLAG-NR2F1-Q244\*, subjected to CRABP2 co-IP with anti-CRABP2 antibody, and immunoblotted with the indicated antibodies. The asterisk indicates the NR2F1/CRABP2 complex. Data representative of three independent experiments. (b) FLAG-NR2F1Q244\* without or with pathogenic mutations was transiently co-transfected alongside E2AziRS and BstYam. Asterisks indicate the NR2F1/CRABP2 complex. Data representative of four independent experiments. (c) Representative NR2F1 (green) and CRABP2 (red) immunostaining of a WT mouse forebrain at E13.5. NR2F1 and CRABP2 co-expression is observed in cells of the hippocampal neuroepithelium (Hpn) (C'–C''). Arrowheads point to representative double-positive cells. Scalebar = 50 μm. GE: ganglionic eminence; Ncx: neocortex; Th: thalamus. (d) NR2F1 (green) and CRABP2 (red) immunostaining of a WT mouse neural retina (Nr) at E13.5. The white arrow in the low magnification image points to high CRABP2 expression in optic nerve (ON) fibers, while arrowheads in high magnification images (D'–D'') point to representative NR2F1/CRABP2 double-positive cells in the ventral Nr. Scalebar = 50 μm. Cr: crystal lens. (e) Top: results from docking simulations of active NR2F1 LBD (bound to 9-cis retinoic acid) with apo- and holo- (bound to retinoic acid) CRABP2. Proteins are represented as surface, NR2F1 LBD is shown in green, apo CRABP2 in yellow, holo CRABP2 in red; retinoic acid and 9-cis retinoic acid are represented in sticks with C atoms in cyan and O atoms in red, the C $\alpha$  of the residues belonging to the interface with CRABP2 are represented as green spheres, those whose mutations are associated with BBSOAS are represented by blue spheres and labeled. Bottom: structural detail of the hydrophobic pockets surrounding the ligands in the active NR2F1 LBD-holo CRABP2 complex. Protein structure is represented as a cartoon with active NR2F1 colored in green and holo CRABP2 in red, ligands are represented as sticks with C atoms in cyan and O atoms in red, the Van der Waals occupancy of each atom belonging to the hydrophobic pockets of either protein is represented as white dots. 90° counterclockwise view of the hydrophobic tunnel potentially allowing the transport of retinoic acid from CRABP2 to NR2F1 LBD.

that the stable pre-formed NR2F1 dimer/oligomer may not be a preferable state in the cellular context. We therefore propose that WT NR2F1 dimerization is a very dynamic, flexible, and transient process, which is in agreement with the common observation that dimeric complexes in signal transduction pathways are generally dynamic, as they act as reversible switches in the process of the information flow (Amoutzias et al., 2008; Nooren & Thornton, 2003). The stable DNA-free NR2F1 homodimer, unbound to DNA, is probably rare and may present disadvantages for the cells as it would prevent the formation of functional complexes between NR2F1 and other partners. Our results corroborate the hypothesis that NR2F1 may instead bind DNA individually and then recruit the second binding partner. Indeed, dimer formation after DNA binding has also been described for other nuclear receptor family members such as RXR, where the key molecular event leading to the cooperative enhancement of dimer formation and DNA binding is the DNA-induced conformational change that creates a favorable interface for protein–protein interactions (Holmbeck et al., 1998). This notion is also supported by studies demonstrating that the sequential monomer-binding pathway allows the protein to search for and locate a specific DNA site more quickly, resulting in greater specificity prior to equilibrium, and thus allowing a single transcription factor to recognize a number of different target sites and fine-tune the activation/repression depending on the dimerizing partners in the cell (Kohler et al., 1999; Kohler & Schepartz, 2001).

### 3.2 | The strength of the genetic code expansion approach

Our data obtained from classical co-IP show that depending on the binding partners, different mutations can have significantly variable effects on the capacity of NR2F1 to form stable dimers/oligomers (Figure 5a–f). We can envisage that pathogenic mutations that favor a higher-order oligomerization of NR2F1 could interfere with the binding of NR2F1 to DNA and other coactivators/repressors and thus inhibit its function. The variation of NR2F1 dimeric structures offers the diversity of target site recognition and function. Therefore, precise information on the impact of pathogenic mutations on dimerization at the cellular level is of great importance. While classical co-IP could provide some information about dimer formation, we could not rule out that the detected complexes included not only dimers but also some higher-order oligomers. Furthermore, the lack of clear detection of WT NR2F1 dimer in the cell hampered the precise evaluation of the effect of pathogenic mutations on dimerization.

Using the site-specific GCE-enabled photocrosslinking of full-length NR2F1 in living cells, we were able to overcome these complications and delineate the differences and similarities among the dimeric species and the pathogenic mutations. Complementary to the *in silico* structural prediction, this novel approach in living cells has allowed us to (i) identify key amino acid residues in contact sites of distinct dimeric pairs, (ii) stably capture transient protein interactions, (iii) demonstrate NR2F1 dimerization at the protein level, (iv) evaluate the effect of BBSOAS mutations on different protein interactions, (v) reveal the proximity of the CRS to the homodimerization and NR2F1/2 dimerization surfaces, and (vi) highlight the diversity of quaternary conformations of full-length NR2F1 in contact with different protein partners.

### 3.3 | The variable functional impact of pathogenic NR2F1 variants

To gain more insight into the relationship between different mutations and variable symptoms of BBSOAS patients, we compared the effects of pathogenic variants with WT NR2F1 in a more functionally relevant manner in living cells. The clear cellular impact of the truncation mutation Q244\*, where a significant portion of the LBD was lost, confirms that the LBD is required for proper nuclear localization and cellular functions of NR2F1 in cell cycle, proliferation, and apoptosis. Interestingly, the G368D mutant, which was predicted to reduce stability and affinity of dimerization while stabilizing the structure of monomeric forms of NR2F1 had no effect on the cellular functions tested. On the other hand, the L372P mutation, predicted to strongly destabilize the monomeric structure, perturbed the cell cycle, reduced cell proliferation, and increased apoptosis similarly to the Q244\* truncation (Figure 3). Finally, the loss of AF2 due to the E400\* truncation significantly hampered cell cycle progression comparably to the Q244\* truncation. Together, the *in cellula* data strongly suggest that the stability of monomeric NR2F1 and the AF2-mediated interaction with co-regulators are a more prominent functional factor compared to the stability and affinity of DNA-free dimers. This is also in agreement with our above-mentioned hypothesis that the stable pre-formed dimerization may be unnecessary for the activity of NR2F1.

### 3.4 | Discovery of CRABP2-NR2F1 binding via the coactivator recognition site

Our data suggest that the dysfunction of NR2F1 in the cellular context correlates well with the disruption of



the monomeric structure integrity due to the L252P mutation in the CRS and the loss of AF2 and thus, the allosteric control on the protein interaction at the CRS due to the E400\* truncation. To date, information about the interacting partners of NR2F1 at the CRS is lacking. The fact that in living human cells NR2F1 heterodimerized with RXR, a key RA-induced co-factor (Evans & Mangelsdorf, 2014), and that NR2Fs either antagonize retinoid-dependent gene expression (Neuman et al., 1995), or are themselves modulated by retinoids (Alfano et al., 2014; Clotman et al., 1998; Tran et al., 1992; Zhuang & Gudas, 2008), suggested us that NR2F1 might be involved in the retinoid signaling pathway during embryonic development by playing distinct functional roles in a context- and time-dependent manner. Moreover, overexpression of some of the NR2F1 variants, in particular, the Q244\* and the L252P mutations, which lead to abnormal localization in the cytoplasm (Figure 4), led us to explore the interaction between NR2F1 and CRABP2, a plasmonuclear shuttling protein, which transports RA to the nucleus and is known to act as a coactivator of RAR (Sessler & Noy, 2005).

Since the CRABP2-RAR complex is a short-lived intermediate (Budhu & Noy, 2002; Dong et al., 1999), we expected that the CRABP2-NR2F1 interaction would have eluded the detection by conventional methods used for studying NR2F1 protein-protein interaction so far. And this was indeed the case. Thanks to the photocrosslinker placed at the CRS (Q244AzF), we could demonstrate for the first time a direct binding of CRABP2 to the CRS of NR2F1 (Figure 8a). Our structural analysis (Figure 8e) also supported our findings and the notion that, in cells, CRABP2 may channel RA to NR2F1 similarly to the RA transfer to RAR (Budhu & Noy, 2002), and that BBSOAS-associated variants may severely perturb the transport of RA. Interestingly, we observed that different pathogenic mutations exerted distinct effects on NR2F1/CRABP2 complex formation. The expression of mutations that strongly destabilize the monomeric structure of NR2F1, namely, the L252P in the CRS and L372P at the RXR $\alpha$ -NR2F1 contact site, abrogated the NR2F1/CRABP2 complex (Figure 8b), again highlighting the monomeric structure integrity of NR2F1 as an important predicting factor of NR2F1 protein-protein interaction. It should also be noted that based on structural alignment with RAR $\beta$  (3KMR), the L252 residue of NR2F1 along with the CRS overlap the predicted binding function-3 (BF-3) of RAR $\beta$ , which is highly conserved among nuclear receptors but has not been extensively characterized (Buzón et al., 2012). Thus, the binding of CRABP2 to CRS/BF-3 of NR2F1 further supports the notion that the functional relationship between the two proteins may be similar to that between CRABP2 and RAR $\beta$ .

On the other hand, the inverse results showing that the CRABP2-NR2F1 binding was strongly enhanced by the G368D mutation (which reduced homo-/heterodimerization) and the G395A mutation (which reduced NR2F1/2 heterodimerization), suggest a scenario where the loss of dimerization, especially with NR2F1 and NR2F2, may shift the balance towards CRABP2-NR2F1 binding and possibly the RA-induced transcription pathway. Since the G395 residue locates between the CRS/BF-3 and the AF2 modules the loss of the allosteric control of these modules likely underlies the aberrantly high CRABP2-NR2F1 interaction. Notably, physiological interactions of NR2F1 with its dimeric partners (NR2F1/2 and RXR $\alpha$ ) *via* DI, and with its putative coactivator (CRABP2) *via* CRS, could be captured by a photocrosslinker placed within the CRS (Q244AzF). These data reveal for the first time the proximity of the DI to the CRS in the quaternary conformation of complexed NR2F1 in living cells, suggesting that, in some cellular contexts, the binding of a dimeric partner at the DI may allosterically hinder the binding of another partner at the CRS. Indeed, the fact that inhibiting the dimerization by mutating the key amino acid residue in DI (G368D) allowed a marked increase in CRABP2 binding at CRS supports this notion.

The fact that most of the pathogenic mutations studied here significantly interfered with the CRABP2 interaction with NR2F1 infers the physiological importance of this interaction. It has been shown that the transcription of NR2F1 and CRABP2 are both induced by RA in mouse embryonic stem cells (Quintero et al., 2018), suggesting context-dependent co-expression of the two proteins. Indeed, our immunofluorescence of the anterior part of the forebrain and the eye region of mouse embryos showed that NR2F1 and CRABP2 are selectively co-expressed in the developing hippocampus and ventral retina, supporting the physiological importance of the CRABP2-mediated NR2F1 function in the brain, particularly in the context of hippocampal and visual pathways. Although apo-CRABP2 is predominantly cytosolic, the RA-bound CRABP2 translocates to the nucleus where it delivers RA to RAR (Budhu & Noy, 2002). In agreement with this known cytoplasmic/nuclear shuttling behavior of CRABP2, the increased CRABP2 binding capacity of the predominantly nuclear-localized NR2F1 variants G368D and G395A suggests that these mutations may favor the interaction between NR2F1 and holo-CRABP2 in the nucleus.

Together, our data demonstrate the allosteric relationship among the conserved sites on the LBD of nuclear receptors, namely the DI, AF2, and CRS/BF-3, along with the consequence of pathogenic mutations therein on the interaction between NR2F1 and its newly identified

partner, CRABP2, and propose the possibility that RA may function as NR2F1 ligand in some physiological contexts.

### 3.5 | Pathogenic mutations, protein interactions, functions, and clinical manifestation of BBSOAS

Drawing from the information we unveiled by structural and *in cellula* analyses about the effects of the pathogenic mutations, we looked for possible connections between the studied NR2F1 variants and the variation in the clinical manifestation of BBSOAS patients. Even if a definitive correlation could not be drawn due to the lack of sufficient clinical data, there are nevertheless some patterns that emerge (Table 2). While all pathogenic mutations examined in the present study are associated with intellectual disability (ID) and most with developmental delay (DD), we can perceive that the additional set of symptoms that include visual impairment (VI), optic atrophy (OA), and optic disc/nerve (OD/ON) anomalies may be more associated with point mutations that strongly destabilized the isolated monomeric structure, reduced NR2F1 protein expression, and promoted stable dimers/oligomers. These mutations (L252P, L372P, and E318D) also caused perturbations in cellular functions investigated in this study, such as nuclear localization defect, cell cycle inhibition, and/or apoptosis. On the contrary, the additional set of symptoms that include epilepsy or seizure, ASD-like symptoms, and motor delay may be more associated with point mutations that reduced NR2F1 homodimer and/or NR2F1/2 or NR2F1-RXR $\alpha$  heterodimer (E318D, G368D, G395A) without significantly perturbing cellular functions investigated in this study. The co-expression of NR2F1 and CRABP2 in the ventral retina and in other brain regions, along with the observation that L252P and L372P mutations (which strongly inhibited CRABP2/NR2F1 binding) are associated with VI, OA, and OD/ON anomalies (while the G368D and G395A mutations are not), infer that the CRABP2-mediated NR2F1 functions may be one of the factors contributing to the visual symptoms of BBSOAS. Future investigation into the role of CRABP2-mediated NR2F1 function is necessary to confirm this hypothesis.

Based on this study, we propose that BBSOAS mutations could be classified into at least four different categories: (i) mutations associated with ID, DD, and visual pathway symptoms (e.g., VI, OA, and OD/ON anomalies), such as L252P and L372P; (ii) mutations associated with ID, DD, epilepsy and ASD symptoms, such as G368D, and G395A; (iii) mutations associated with wide range of symptoms such as E318D and E400\*; and

(iv) mutations associated with ID, DD without visual pathway, epilepsy or ASD symptoms, such as the Q244\* truncation. The fact that the Q244\* mutation, which was excluded from the nucleus and exerted the strongest perturbation on cellular functions, does not cause the additional visual pathway, epilepsy, or ASD symptoms, suggests the possibility that ID, DD, and motor delay may be associated with loss of nuclear NR2F1 function or perturbation of the function/transport of other protein partners in the cytoplasm. Other additional symptoms may also be attributed to modifications in the nuclear functions or NR2F1 variants, such as changes in protein–protein interactions (dimerization partners, coactivators, corepressors), which might also act as dominant negative perturbators. The E400\* mutation, which causes the loss of AF2, the allosteric control of coactivator/corepressor binding and the partial exclusion from the nucleus, is associated not only with ID, DD, and motor delay, but also with other symptoms, thus supporting this view. As an additional note, the E318D mutation, which is not located in any previously described domain or interaction site but was predicted to destabilize the monomeric NR2F1, reduced NR2F1 expression and caused NR2F1, NR2F1/2 stable dimers/oligomers, is linked to a wide range of symptoms. Such observations indicate the importance of this site, which warrants more investigation. Further studies based on disease models and more clinical data will be needed to confirm our initial deductions and provide further genotype-phenotype correlations of BBSOAS.

## 4 | CONCLUSION

One of the crucial pieces of information to help clinicians improve diagnostic accuracy, treatment efficacy, and predict their response to different interventions is to draw a clear genotype–phenotype correlation among patients. This requires knowledge of the structure–function relationship of the pathogenic protein. In this study, we combined computational analysis of protein structure and genetic code expansion technology to deliver new insight into the structure–function relationship of NR2F1 in the context of BBSOAS pathology. The structural analysis of the isolated and dimeric LBD has provided an understanding of the molecular interaction between this domain and its partners without the complex interactions from other cellular partners. On the other side, the complementary GCE-enabled site-specific photocrosslinking in living cells has highlighted the variable quaternary conformations of NR2F1 with functional relevance, where the contribution of different domains of the full-length protein and interaction with other cellular actors





were also taken into account. Thanks to this dual approach, our data have contributed to associate the variable BBSOAS symptoms to different point mutations in the NR2F1 LBD. Although HEK293T cells are not the most clinically relevant cells for studying neurodevelopmental defects, we utilized them to allow us to focus on dissecting basic biological processes, such as cell division and cell death, and protein interactions at the molecular level. More neurodevelopmentally oriented studies and thorough identification of NR2F1 co-factors will be needed to be performed in a neural-specific cell environment. We believe that the complementary approaches described here not only provide information that will contribute to a better understanding of the genotype-phenotype correlation in BBSOAS patients but can also be applied to a wide range of investigations involving protein structure-function relationships in other genetic diseases.

## 5 | MATERIALS AND METHODS

### 5.1 | Molecular modeling of NR2F1 ligand-binding domain (LBD) in its auto-repressed and active conformations

The three-dimensional structure of human NR2F1 LBD in its auto-repressed conformation was obtained by homology modeling using the crystal structure of human NR2F2 LBD as a template (PDB entry 3CJW, resolution 1.48 Å° (Kruse et al., 2008), which shares 96% sequence identity with that of NR2F1. The missing loops encompassing residues 201–213 and 276–292 were modeled using Maestro (Schroedinger). The structure of human NR2F1 LBD in its active conformation was modeled as previously elucidated in (Khalil et al., 2022), thus using the structure of RXRα in complex with 9-cis-retinoic acid (9cRA) as a template (PDB entry 1FM6 [Gampe Jr. et al., 2000], ~40% sequence identity with human NR2F1 LBD) after removal of the coactivator peptides. The resulting model of active NR2F1 LBD was then employed for docking and mutagenesis analyses. All molecular models were prepared following the *Protein Preparation Wizard* pipeline included in Bioluminate (Schroedinger), which provides for assigning bond orders according to the Chemical Components Dictionary database ([www.pdb.org](http://www.pdb.org), wwPDB Foundation, Piscataway, NJ), adding H atoms, selecting the most probable rotamer in case of alternative conformations of the sidechains and modeling of the missing loops. The structures were then refined by sampling the orientation of water molecules and predicting the protonation states of ionizable amino acids at pH 7.5 using PROPKA (Li et al., 2005) to assign and

optimize H-bonds. The last step of protein preparation consisted of the minimization of the model structures using OPLS4 forcefield (Schroedinger, New York, NY) using 0.3 Å as threshold for the Root-Mean Square Displacement (RMSD) of heavy atoms. Regarding *in silico* docking simulations, the three-dimensional structural model of human NR2F2 LBD was obtained by applying the *Protein Preparation Wizard* pipeline to the PDB entry 3CJW (Kruse et al., 2008), where missing residues 194–206 and 269–285 were modeled by Maestro (Schroedinger). Human RXRα was modeled by following the *Protein Preparation Wizard* pipeline for the respective molecule present in PDB entry 4NQA. The PDB files of the modeled NR2F1 structures (in both the auto-repressed and active form) are available at <https://zenodo.org/records/10551664>.

### 5.2 | Docking simulations of NR2F1 LBD in its auto-repressed and active state

To establish a putative NR2F1 interaction model in native-like conditions, we superimposed the experimental structures of the dimeric assemblies of homologous nuclear receptors, namely NR2F2 homodimer (PDB entry 3CJW) (Kruse et al., 2008), RXRα homodimer (PDB entry 1MZN) (Egea et al., 2002), RXRα-Retinoic Acid Receptor β (RARβ) heterodimer (PDB entry 5UAN) (Chandra et al., 2017) and RXRα-LXRβ heterodimer (PDB entry 4NQA) (Lou et al., 2014). All experimental dimers presented highly similar and conserved (Figure S4) interfaces in terms of the physicochemical properties of interacting residues, therefore the dimeric assembly corresponding to NR2F2 homodimer was selected as the native-like conformation. Rigid-body docking simulations of NR2F1 LBD with its putative partners, namely NR2F1 (homodimer), NR2F2 and RXRα LBDs, were performed using PIPER (Kozakov et al., 2006) module implemented in Bioluminate (Schroedinger), by setting 70,000 as the number of tested poses with a 5° sampling of Euler angles. Modeled loops 201–213, 276–292 and the C-terminal loop (residues 407–414) of both auto-repressed and active NR2F1 LBD were excluded from the docking interface to avoid potential artifacts due to highly flexible regions. For the same reason, modeled loops 193–207, 268–286, and 400–407 of human NR2F2 LBD were masked during docking simulations. The best 1000 resulting poses from docking simulations were then grouped in ~100 clusters, whose centroids were aligned against the native-like conformation to identify the native-like cluster with the lowest RMSD with respect to PDB entry: 3CJW (NR2F2 homodimer). Finally, the highest-scored conformation of the

native-like cluster was selected as the model for WT NR2F1 LBD-partner interaction. Each of the eight obtained complexes (NR2F1 homodimer in homo- and heterozygosis, NR2F1-NR2F2 heterodimer and NR2F1-RXR $\alpha$  heterodimer, with NR2F1 in both its auto-repressed and active form) underwent the *Protein Preparation Wizard* pipeline prior to *in silico* mutagenesis with the same parameters as detailed in the previous section. The modeled NR2F1 LBD was also subjected to rigid-body docking using ZDOCK 2.3 (Chen et al., 2003) to evaluate homodimer formation propensity based on the correlation between ZDOCK score and experimentally determined binding constants (Dell'Orco et al., 2007). We carried out four rigid-body docking simulations for both the auto-repressed and the active form of NR2F1 LBD, starting from four different relative orientations of the monomers with dense sampling (6° sampling step), yielding 4000 potential docking poses for each simulation. The final 16,000 poses were filtered to identify those resembling the NR2F1 dimer obtained with PIPER, by setting the threshold of the C $\alpha$ -RMSD to 1 Å. The PDB files of the docked NR2F1 (in both the auto-repressed and active form) homodimers and heterodimers with NR2F2 and RXR $\alpha$  can be found at <https://zenodo.org/records/10551664>.

### 5.3 | Molecular modeling of CRABP2 and docking simulations with active NR2F1 LBD

The structures of apo and holo human CRABP2 (unbound/bound to retinoic acid) were retrieved from PDB entries 2FS6 and 2FS3, respectively (Vaezeslami et al., 2006). Before docking simulations, both CRABP2 models underwent the same *Protein Preparation Wizard* procedure as elucidated in the previous sections. Unbiased docking simulations of active NR2F1-LBD with CRABP2 were carried out using PIPER (Kozakov et al., 2006) module implemented in Bioluminate (Schroedinger), by setting 70,000 as the number of tested poses with a 5° sampling of Euler angles. The best 1000 resulting poses from docking simulations were then grouped in ~30 clusters, whose centroids were filtered according to the simultaneous presence of CRABP2 residues Q75, P81, and K102 (whose mutations compromise binding of CRABP2 to RXR $\alpha$ ) (Budhu & Noy, 2002) and NR2F1 residue Q244 (used for GCE crosslinking) in the NR2F1-CRABP2 interface. Finally, the C $\alpha$  of the four filtered solutions (2 for apo and 2 for holo CRABP2) obtained from docking simulations were superimposed on each other to identify the potential native-like conformation (RMSD = 2.123 Å) in which apo and holo

CRABP2 essentially shared the interaction interface with NR2F1 regardless of the presence of the retinoid. The PDB files of the docked active NR2F1-CRABP2 complex (in both the apo and holo form) can be found at <https://zenodo.org/records/10551664>.

### 5.4 | In silico prediction of relative stabilities and affinities of NR2F1 LBD missense variants

All missense mutations in NR2F1 LBD monomer, homo- and heterodimers were introduced using the *Residue Scanning* tool from Bioluminate (Schroedinger); for NR2F1 homodimer mutations were generated in both homo- and heterozygosis. The most probable rotamer for the sidechain of each aminoacidic substitution was automatically assigned before performing the same energy minimization protocol as above. We estimated the effects of each substitution on the Gibbs free energy of folding ( $\Delta\Delta G_f^{\text{app}} = \Delta G_f^{\text{app,mut}} - \Delta G_f^{\text{app,WT}}$ ) of both the auto-repressed and the active forms of isolated NR2F1 LBD and its complex with NR2F1 (both in homo- and heterozygosis), NR2F2 and RXR $\alpha$  compared to the WT. The computation of  $\Delta\Delta G_f^{\text{app}}$ , expressed in kcal/mol, was performed by calculating the variant-specific thermodynamic cycle using the Molecular Mechanics/Generalized Born and Surface Area Continuum solvation (MM/GBSA) method, which does not include the explicit energetic contribution deriving from conformational changes. Therefore, the  $\Delta\Delta G_f^{\text{app}}$  reported in Tables S1–S3, should not be considered as precise thermodynamic quantities, but rather “apparent” values. Positive values of  $\Delta\Delta G_f^{\text{app}}$  indicate a destabilizing effect for the specific mutation, whereas negative values identify a stabilizing mutation. The variations in Gibbs free energy of binding ( $\Delta\Delta G_b^{\text{app}} = \Delta G_b^{\text{app,mut}} - \Delta G_b^{\text{app,WT}}$ ) reported in Tables S2 and S3 were calculated between NR2F1-LBD and the partner nuclear receptor (NR2F1 in homo/heterozygosis, NR2F2 and RXR $\alpha$ ). Positive values of  $\Delta\Delta G_b^{\text{app}}$  indicate decreased affinity for the specific ligand, whereas negative values identify stronger binding.

### 5.5 | Molecular dynamics simulations of NR2F1-LBD nonsense mutations

Molecular Dynamics simulations of isolated human NR2F1-LBD variants were set up based on the model of the auto-repressed form described in the previous sections. The models for nonsense variants E400\* and Q244\* were generated by truncating the structure of the WT after the carbonyl group at the C-term of the respective

residue and capping both N- and C-terms with a  $\text{NH}_2$  group. All protein systems underwent the same *Protein Preparation Wizard* procedure described in previous sections before setting up the simulation system. All-atom MD simulations of NR2F1-LBD variants were run on GROMACS 2016.5 (Abrahams et al., 2013) simulation package, setting CHARMM36m (Huang et al., 2017) as forcefield. The size of the simulated protein systems was  $\sim 45,000$  atoms for WT and E400\* variants and  $\sim 30,000$  atoms for Q244\* mutant, systems were prepared and minimized according to the protocol and parameters detailed in (Marino et al., 2015). Briefly, proteins were put in the center of a dodecahedral box with the edges set 12 Å apart from any protein atom to avoid potential interactions with the periodic images, then the systems were neutralized with 150 mM KCl and underwent a two-step energy minimization procedure with steepest descent and conjugate gradients algorithms. System equilibration was carried out for 2 ns in NVT ensemble ( $T = 310$  K) with position restraints followed by 2 ns with no position restraints as explained in (Marino & Dell'Orco, 2016), while productive 500 ns runs were carried out in NPT ensemble ( $T = 310$  K,  $P = 1$  atm) using the setup elucidated in (Marino & Dell'Orco, 2019). The  $\text{C}\alpha$ -RMSD with respect to the protein after equilibration was calculated using *gmx rms*, while the time-averaged  $\text{C}\alpha$ -RMSD over the 500 ns trajectory with respect to the average structure, that is the Root-Mean Square Fluctuation (RMSF), was calculated by *gmx rmsf*. To evaluate the conformational rearrangement due to the chain truncations, we calculated the angles between helices 1 (identified by the vector connecting the  $\text{C}\alpha$  of C183 and R194) and 2 (identified by the vector connecting  $\text{C}\alpha$  of C220 and R235) and between helices 1 and 8 (identified by the vector connecting  $\text{C}\alpha$  of S382 and F390) using *gmx gangle*. All *gmx* functions were provided by GROMACS 2016.5 simulation package. The starting, equilibrated, and final structures of the modeled NR2F1 nonsense variants, as well as the simulations in GROMACS compressed format (\*.xtc) can be found at <https://zenodo.org/records/10551664>.

## 5.6 | Cell preparation and transfection

The HEK293T cell line was maintained at 37°C in a humidified incubator with 5%  $\text{CO}_2$  in DMEM supplemented with 10% fetal bovine serum (FBS). The cells were seeded at 350,000 cells per well in 6-well plates for oligomerization and dimerization experiments, 150,000 cells per well in 12-well plates to study the cell cycle and apoptosis, and 80,000 cells per well in 24-well plates for immunofluorescent staining. After 24 h, the cells were

transfected with JetPRIME transfection reagent with the 1.5, 1, and 0.1875  $\mu\text{g}$  total DNA amount for 6-, 12-, and 24-well plates respectively. For experiments using GCE, co-transfection of pIRE4-Azi and pcDNA3.1-NR2F1 DYK was done using 3:1 ratio (1.125:0.375  $\mu\text{g}/\mu\text{l}$ ) in 6-well plate, respectively. In case of co-transfection with three plasmids using GCE, pIRE4-Azi, pcDNA3.1-NR2F1 DYK, and protein partners-Myc tag were used as 3:1:1 ratio (0.9:0.3:0.3  $\mu\text{g}/\mu\text{l}$ ) in 6-well plate.

## 5.7 | Plasmids

The plasmid for the expression of the C-terminally FLAG-tagged human NR2F1, pcDNA3.1-NR2F1-DYK (Genscript clone ID: OHu23866D), was used as the template for the site-directed mutagenesis (Liu & Naismith, 2008) to produce NR2F1 variants with amber codon and pathogenic mutations. To generate the constructs to express Myc-tagged protein partners, that is, pcDNA-Myc-NR2F1, pcDNA-Myc-NR2F2, and pcDNA-Myc-RXR $\alpha$ , EcoRI restriction site was created by site-directed mutagenesis upstream of the NR2F1, NR2F2, and RXR $\alpha$  genes in pcDNA3.1-NR2F1-FLAG, pME-NR2F2, and pcDNA3.1-hRAR $\alpha$ .hRXR $\alpha$ , respectively. The genes were then excised and cloned to replace DBC1 in the pcDNA Myc DBC1 plasmid by restriction ligation at the EcoRI/XhoI restriction sites. pME-NR2F2 was a gift from Nathan Lawson (Addgene plasmid #138359; <http://n2t.net/addgene:138359>; RRID:Addgene\_138359). pcDNA Myc DBC1 was a gift from Osamu Hiraike (Addgene plasmid #35096 <http://n2t.net/addgene:35096>; RRID:Addgene\_35096). pcDNA3.1-hRAR $\alpha$ .hRXR $\alpha$  was a gift from Catharine Ross (Addgene plasmid #135411; <http://n2t.net/addgene:135411>; RRID:Addgene\_135411). pIRE4-Azi was a kind gift from Irene Coin (Addgene plasmid #105829; <http://n2t.net/addgene:105829>; RRID:Addgene\_105829). All plasmid used in this study are listed in Table S5.

## 5.8 | Incorporation of AzF and photo-crosslinking

Twenty-four hours before the transfection, cells were seeded at 350,000 cells per well in 6-well plates. Using the JetPRIME transfection reagent (Polyplus), cells were co-transfected with pIRE4-AziRS and pcDNA3.1-NR2F1-FLAG (WT or variants) plasmids at the ratio of 3:1, respectively. For dimerization experiments, the plasmids carrying Myc-NR2F1, NR2F2, or RXR $\alpha$  were transfected at the same amount as FLAG-NR2F1. A fresh 150-mM p-azido-l-phenylalanine (AzF) solution was prepared on the day of the transfection by dissolving the AzF powder (Santa Cruz) in 0.375 N



NaOH and 25% DMSO. A 3-mM working solution (WS) of AzF in a complete medium supplemented with 100 mM HEPES was prepared and added to the well to obtain the final AzF concentration of 300  $\mu$ M. Forty-eight hours after the transfection, the cells washed twice with DMEM to remove AzF and the medium was replaced with the DMEM containing 10% FBS and 10 mM HEPES. Protein photocrosslinking was performed by irradiating the cells with 365-nm UV for 15 min at room temperature under a UV lamp (UVITEC LF-215.LS LAMP 365/254NM 2X15W 230V).

## 5.9 | Immunoprecipitation

Cells were washed with PBS and lysed with prechilled RIPA buffer (50 mM Tris-HCL, pH 8.0 with 150 mM sodium chloride, 1.0% NP-40, 0.5% sodium deoxycholate and 0.1% sodium dodecyl sulfate) with cocktail protease inhibitor (cOmplete, Sigma-aldrich) for 30 min in at 4°C on a shaker. Lysed cells were scraped off the plate and transferred to 1.5  $\mu$ L Eppendorf tube and sonicated for 5 s on ice. Cells debris were removed by centrifugation (14,000 $\times$ g, 10 min, 4°C). The supernatants were immunoprecipitated using ANTI-FLAG M2 Affinity Gel (Sigma-Aldrich) or Sepharose beads coupled with the anti-CRABP2 antibody (Proteintech) overnight at 4°C. The beads were collected by centrifugation (5000 $\times$ g, 1 min, 4°C) and washed three times with TBS. The immunoprecipitated proteins were eluted from the beads in LDS sample buffer (NuPAGE LDS Sample Buffer, NP0008, ThermoFisher) by heating for 10 min at 70°C. The eluted proteins were separated from the beads by centrifugation (5000 $\times$ g, 1 min) and transferred to 1.5  $\mu$ L Eppendorf tube. Proteins were denatured by dithiothreitol at the final concentration 0.1 M. All antibodies used in this study are listed in Table S6.

## 5.10 | Western blot

The protein samples were electrophoresed in the 6%–15% gradient polyacrylamide gels and blotted onto the PVDF membrane (Amersham Hybond P 0.45 PVDF blotting membrane, Cytiva). The proteins of interest were probed using the primary antibodies, that is, mouse anti-Myc tag (Cell Signaling Technology), rabbit anti-DDDDK (FLAG) tag (Genetex), rabbit anti-histone H3 (R&D), rabbit anti-CRABP2 (Proteintech), and peroxidase-conjugated anti-rabbit or anti-mouse secondary antibodies. The proteins were then detected based on chemiluminescence (ECL Prime Western Blotting System, Cytiva) using Fusion FX7 imager (Vilber Lourmat). Regarding the

densitometric analysis, immunoblot band intensities were quantified using the ImageJ software (Schneider et al., 2012). For the analysis of NR2F1 interaction with dimeric partners (NR2F1, NR2F2, and RXR) by conventional IP and GCE-enabled photocrosslinking IP, the band intensity of the co-IP Myc-tagged partner was divided by the intensity of each FLAG-NR2F1 variant (WT or mutants) ( $R_1 = \text{Myc-partner}/\text{FLAG-NR2F1}$ ). A final normalized intensity value ( $R_{\text{fin}}$ ) presented in the plot was obtained by normalizing  $R_1$  of from each NR2F1 variant ( $R_{1\text{var}}$ ) against  $R_1$  from wild-type NR2F1 ( $R_{1\text{wt}}$ ) ( $R_{\text{fin}} = R_{1\text{var}}/R_{1\text{wt}}$ ,  $R_{\text{fin}} = 1$  for NR2F1 WT).

## 5.11 | Immunofluorescence microscopy

After 48 h of transfection, HEK293T cells were fixed with 4% paraformaldehyde (PFA) for 15 min at room temperature (RT) and then washed 3 times with PBS. The cells were pre-blocked with pre-blocking solution (PBS, 0.3% tween, 5% serum) for 1 h at RT with gentle shaking every 15 min. Cells were stained with primary antibodies; rabbit anti-NR2F1 (1:1000) (Abcam) and mouse anti-acetylated-tubulin (1:1000) (Sigma) for 4 h at RT with gentle shaking every hour, followed by washing with PBS containing 1% serum. After washing, cells were stained with secondary antibodies (Alexa Fluor-488 anti-rabbit and Alexa Fluor-647 anti-mouse antibodies) and DAPI for 1 h at RT and washed 3 times with PBS. Cells were mounted on glass coverslips with mounting medium (PBS, 2% N-propyl gallate, 90% Glycerol) and analyzed using Zeiss 710 confocal microscope equipped with a 405 nm diode, an argon ion, a 561 nm DPSS, and 647 HeNe lasers using a  $\times$ 40 objective with oil immersion. Confocal images were obtained with single plane acquisitions; cell counting was performed manually on ImageJ. All antibodies used in this study are listed in Table S6.

## 5.12 | Immunostaining and FACS for cell cycle and apoptosis assays

After 48 h of transfection with pcDNA-NR2F1-FLAG (WT or pathogenic variants), HEK293T cells were fixed with 70% cold ethanol for 15 min and then washed with 3 mL of PBS containing 1% FBS. The fixed cells were blocked with a pre-blocking solution (PBS, 0.3% tween, 5% serum) and then incubated for 30 min at 4°C with gentle shaking. The cells were co-stained with primary antibodies, mouse anti-NR2F1 (1:1000) (R&D), rabbit anti-phospho-Histone H3 (pH 3) (1:1000) (Millipore), or rabbit anti-cleaved Caspase 3 (1:1000) (Cell signaling)

and incubated for 2 h at 4°C with gentle shaking. After washing, cells were stained with Alexa Fluor 488 anti-mouse and Alexa Fluor 647 anti-rabbit antibodies for 1 h at 4°C with gentle shaking. Then the cells were washed with PBS. To study the cell cycle, cells were stained with propidium iodide (PBS, 0.1% NP40, 0.2% RNase, 3.95% propidium iodide) for 15 min at RT followed by two washes with PBS. After filtering through a 70-micron filter, the cells were analyzed by FACS using a BD LSRFortessa and FACSDiva software (Becton Dickinson). Cells were analyzed on the basis of 10,000 total events (debris excluded) per experiment; for cell cycle and apoptosis analyses, cell percentages were calculated over the number of NR2F1+ cells and normalized on the control sample (cells transfected with full-length WT NR2F1). All antibodies used are listed in Table S6.

### 5.13 | Immunofluorescence of mouse brain

All mouse experiments were conducted in accordance with the relevant national and international guidelines and regulations (European Union rules; 2010/63/UE), and with approval by the local ethical committee in France (CIEPAL NCE/2019–548). Standard housing conditions were approved by the local ethical committee in France (CIEPAL NCE/2019–548). Briefly, adult mice were kept on a 12-h light–dark cycle and three animals were housed per cage with the recommended environmental enrichment (wooden cubes, cotton pad and igloo), and with food and water *ad libitum*. Whole heads of embryonic (E) 13.5 mouse embryos were dissected on ice-cold PBS 1× and fixed in 4% paraformaldehyde (PFA) at 4°C for 3 h in agitation, then washed in PBS 1× and dehydrated in 25% sucrose overnight at 4°C. Heads were then embedded in optimal cutting temperature compound (OCT) and stored at –80°C. Cryostat sections (12–14 μm) were collected on SuperFrost slides and subjected to immunofluorescence, as previously described (Harb et al., 2021). Briefly, the brain sections were washed and unmasked in sodium citrate 85 mM (pH 6), 95–100°C for 10 min. After pre-blocking, the samples were stained with rabbit anti-CRABP2 (102251-AP; Proteintech) and mouse anti-NR2F1 (H8132; R&D) antibodies followed by corresponding secondary antibodies and counterstaining with DAPI. After subsequent washing, sections were covered with mounting solution (80% glycerol, 2% N-propyl gallate in PBS 1×) and glass coverslips. Images were acquired using an Apotome Zeiss microscope with a ×20 objective using the AxioVision software, and exported as TIF files, then opened in Photoshop or ImageJ for further analysis. All antibodies used are listed in Table S6.

### 5.14 | Statistical analysis

Statistical analyses (ANOVA or *t*-test) were performed using GraphPad Prism9 and data are presented as mean ± SD. For 2-way ANOVA, Dunnett's multiple comparison was performed as a post hoc analysis. N values represent biological replicates from at least three independent experiments, unless otherwise stated in figure legends.

#### AUTHOR CONTRIBUTIONS

**Michèle Studer:** Conceptualization; investigation; funding acquisition; writing – original draft; writing – review and editing; supervision; project administration; resources. **Valerio Marino:** Methodology; validation; formal analysis; software; investigation; data curation; resources; conceptualization; visualization; writing – review and editing. **Wanchana Phromkrasae:** Conceptualization; investigation; methodology; validation; formal analysis; data curation; visualization; writing – review and editing. **Michele Bertacchi:** Investigation; visualization; validation; methodology; writing – review and editing; data curation. **Paul Cassini:** Investigation; validation; visualization. **Krittalak Chakrabandhu:** Conceptualization; writing – original draft; writing – review and editing; supervision; methodology; data curation. **Daniele Dell'Orco:** Conceptualization; investigation; methodology; software; data curation; supervision; formal analysis; funding acquisition; writing – original draft; writing – review and editing.

#### ACKNOWLEDGMENTS

We thank the Genetic Code Expansion facility at iBV for the assistance in applying the GCE technology. We also thank A. Loubat at the flow cytometry facility at iBV. This study was funded by the French Government (National Research Agency, ANR) through the ‘Investments for the Future’ programs IDEX UCAJedi ANR-15-IDEX-01, by the “Fondation de la Recherche Médicale (Equipe FRM2020)” (#EQU202003010222), “Fondation de France” (#00123416), ERA-NET Neuron grant (Brain4Sight) (ANR-21-NEU2-0003-03) grants to Michèle Studer. Valerio Marino was the recipient of a research contract within the FSE REACT EU-PON R&I 2014–2020 granted to Daniele Dell'Orco. Partial funding was also obtained by the Next Generation EU/Ministry of University and Research project: “A multiscale integrated approach to the study of the nervous system in health and disease (MNESYS)”, CUP B33C22001060002, PE00000006 missione 4, componente 2, investimento 1.3.

#### CONFLICT OF INTEREST STATEMENT

The authors declare no financial and non-financial competing interests.

## ORCID

Michèle Studer  <https://orcid.org/0000-0001-7105-2957>

## REFERENCES

- Abrahams BS, Arking DE, Campbell DB, Mefford HC, Morrow EM, Weiss LA, et al. SFARI gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). *Mol Autism*. 2013;4:36.
- Alfano C, Magrinelli E, Harb K, Studer M. The nuclear receptors COUP-TF: a long-lasting experience in forebrain assembly. *Cell Mol Life Sci*. 2014;71:43–62.
- Amoutzias GD, Pichler EE, Mian N, de Graaf D, Imsiridou A, Robinson-Rechavi M, et al. A protein interaction atlas for the nuclear receptors: properties and quality of a hub-based dimerization network. *BMC Syst Biol*. 2007;1:34.
- Amoutzias GD, Robertson DL, Van de Peer Y, Oliver SG. Choose your partners: dimerization in eukaryotic transcription factors. *Trends Biochem Sci*. 2008;33:220–9.
- Bertacchi M, Parisot J, Studer M. The pleiotropic transcriptional regulator COUP-TFI plays multiple roles in neural development and disease. *Brain Res*. 2019;1705:75–94.
- Bertacchi M, Romano AL, Loubat A, Tran Mau-Them F, Willems M, Faivre L, et al. NR2F1 regulates regional progenitor dynamics in the mouse neocortex and cortical gyrification in BBSOAS patients. *EMBO J*. 2020;39:e104163.
- Bertacchi M, Tocco C, Schaaf CP, Studer M. Pathophysiological heterogeneity of the BBSOA neurodevelopmental syndrome. *Cells*. 2022;11:1260.
- Billiet B, Amati-Bonneau P, Desquiret-Dumas V, Guehlouz K, Milea D, Gohier P, et al. NR2F1 database: 112 variants and 84 patients support refining the clinical synopsis of Bosch-Boonstra-Schaaf optic atrophy syndrome. *Hum Mutat*. 2022;43:128–42.
- Bosch DG, Boonstra FN, Gonzaga-Jauregui C, Xu M, de Ligt J, Jhangiani S, et al. NR2F1 mutations cause optic atrophy with intellectual disability. *Am J Hum Genet*. 2014;94:303–9.
- Budhu AS, Noy N. Direct channeling of retinoic acid between cellular retinoic acid-binding protein II and retinoic acid receptor sensitizes mammary carcinoma cells to retinoic acid-induced growth arrest. *Mol Cell Biol*. 2002;22:2632–41.
- Capriotti E, Fariselli P, Calabrese R, Casadio R. Predicting protein stability changes from sequences using support vector machines. *Bioinformatics*. 2005;21(Suppl 2):ii54–8.
- Capriotti E, Fariselli P, Casadio R. A neural-network-based method for predicting protein stability changes upon single point mutations. *Bioinformatics*. 2004;20(Suppl 1):i63–8.
- Chandra V, Wu D, Li S, Potluri N, Kim Y, Rastinejad F. The quaternary architecture of RARbeta-RXRalpha heterodimer facilitates domain-domain signal transmission. *Nat Commun*. 2017;8:868.
- Chen CA, Bosch DG, Cho MT, Rosenfeld JA, Shinawi M, Lewis RA, et al. The expanding clinical phenotype of Bosch-Boonstra-Schaaf optic atrophy syndrome: 20 new cases and possible genotype-phenotype correlations. *Genet Med*. 2016;18:1143–50.
- Chen R, Li L, Weng Z. ZDOCK: an initial-stage protein-docking algorithm. *Proteins*. 2003;52:80–7.
- Clotman F, Van Maele-Fabry G, Picard JJ. All-trans-retinoic acid upregulates the expression of COUP-TFI in early-somite mouse embryos cultured in vitro. *Neurotoxicol Teratol*. 1998;20:591–9.
- Cooney AJ, Tsai SY, O'Malley BW, Tsai MJ. Chicken ovalbumin upstream promoter transcription factor (COUP-TF) dimers bind to different GGTC A response elements, allowing COUP-TF to repress hormonal induction of the vitamin D3, thyroid hormone, and retinoic acid receptors. *Mol Cell Biol*. 1992;12:4153–63.
- Dell'Orco D. Fast predictions of thermodynamics and kinetics of protein-protein recognition from structures: from molecular design to systems biology. *Mol Biosyst*. 2009;5:323–34.
- Dell'Orco D, De Benedetti PG, Fanelli F. In silico screening of mutational effects on enzyme-protein inhibitor affinity: a docking-based approach. *BMC Struct Biol*. 2007;7:37.
- Dobson CM. Protein folding and misfolding. *Nature*. 2003;426:884–90.
- Dong D, Ruuska SE, Levinthal DJ, Noy N. Distinct roles for cellular retinoic acid-binding proteins I and II in regulating signaling by retinoic acid. *J Biol Chem*. 1999;274:23695–8.
- Egea PF, Mitschler A, Moras D. Molecular recognition of agonist ligands by RXRs. *Mol Endocrinol*. 2002;16:987–97.
- Evans RM, Mangelsdorf DJ. Nuclear receptors, RXR, and the big bang. *Cell*. 2014;157:255–66.
- Fokkema I, van der Velde KJ, Slofstra MK, Ruivenkamp CAL, Vogel MJ, Pfundt R, et al. Dutch genome diagnostic laboratories accelerated and improved variant interpretation and increased accuracy by sharing data. *Hum Mutat*. 2019;40:2230–8.
- Gampe RT Jr, Montana VG, Lambert MH, Miller AB, Bledsoe RK, Milburn MV, et al. Asymmetry in the PPARgamma/RXRalpha crystal structure reveals the molecular basis of heterodimerization among nuclear receptors. *Mol Cell*. 2000;5:545–55.
- Germain P, Staels B, Dacquet C, Spedding M, Laudet V. Overview of nomenclature of nuclear receptors. *Pharmacol Rev*. 2006;58:685–704.
- Harb K, Bertacchi M, Studer M. Optimized immunostaining of embryonic and early postnatal mouse brain sections. *Bio Protoc*. 2021;11:e3868.
- Holmbeck SM, Dyson HJ, Wright PE. DNA-induced conformational changes are the basis for cooperative dimerization by the DNA binding domain of the retinoid X receptor. *J Mol Biol*. 1998;284:533–9.
- Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, de Groot BL, et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat Methods*. 2017;14:71–3.
- Jezela-Stanek A, Ciara E, Jurkiewicz D, Kucharczyk M, Jedrzejowska M, Chrzanowska KH, et al. The phenotype-driven computational analysis yields clinical diagnosis for patients with atypical manifestations of known intellectual disability syndromes. *Mol Genet Genomic Med*. 2020;8:e1263.
- Jurkute N, Bertacchi M, Arno G, Tocco C, Kim US, Kruszewski AM, et al. Pathogenic NR2F1 variants cause a developmental ocular phenotype recapitulated in a mutant mouse model. *Brain Commun*. 2021;3:fcab162.
- Kaiwar C, Zimmermann MT, Ferber MJ, Niu Z, Urrutia RA, Klee EW, et al. Novel NR2F1 variants likely disrupt DNA binding: molecular modeling in two cases, review of published cases, genotype-phenotype correlation, and phenotypic expansion of the Bosch-Boonstra-Schaaf optic atrophy syndrome. *Cold Spring Harbor Mol Case Stud*. 2017;3:a002162.
- Khalil BD, Sanchez R, Rahman T, Rodriguez-Tirado C, Moritsch S, Martinez AR, et al. An NR2F1-specific agonist suppresses metastasis by inducing cancer cell dormancy. *J Exp Med*. 2022;219:e20210836.

- Klinge CM, Silver BF, Driscoll MD, Sathya G, Bambara RA, Hilf R. Chicken ovalbumin upstream promoter-transcription factor interacts with estrogen receptor, binds to estrogen response elements and half-sites, and inhibits estrogen-induced gene expression. *J Biol Chem*. 1997;272:31465–74.
- Kohler JJ, Metallo SJ, Schneider TL, Schepartz A. DNA specificity enhanced by sequential binding of protein monomers. *Proc Natl Acad Sci U S A*. 1999;96:11735–9.
- Kohler JJ, Schepartz A. Kinetic studies of Fos-Jun.DNA complex formation: DNA binding prior to dimerization. *Biochemistry*. 2001;40:130–42.
- Kozakov D, Brenke R, Comeau SR, Vajda S. PIPER: an FFT-based protein docking program with pairwise potentials. *Proteins*. 2006;65:392–406.
- Kruse SW, Suino-Powell K, Zhou XE, Kretschman JE, Reynolds R, Vonrhein C, et al. Identification of COUP-TFII orphan nuclear receptor as a retinoic acid-activated receptor. *PLoS Biol*. 2008;6:e227.
- Leng X, Cooney AJ, Tsai SY, Tsai MJ. Molecular mechanisms of COUP-TF-mediated transcriptional repression: evidence for transrepression and active repression. *Mol Cell Biol*. 1996;16:2332–40.
- Li H, Robertson AD, Jensen JH. Very fast empirical prediction and rationalization of protein pKa values. *Proteins*. 2005;61:704–21.
- Liu H, Naismith JH. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol*. 2008;8:91.
- Lou X, Toresson G, Benod C, Suh JH, Philips KJ, Webb P, et al. Structure of the retinoid X receptor alpha-liver X receptor beta (RXRalpha-LXRbeta) heterodimer on DNA. *Nat Struct Mol Biol*. 2014;21:277–81.
- Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, et al. CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res*. 2020;48:D265–8.
- Marino V, Dell'Orco D. Allosteric communication pathways routed by Ca(2+)/Mg(2+) exchange in GCAP1 selectively switch target regulation modes. *Sci Rep*. 2016;6:34277.
- Marino V, Dell'Orco D. Evolutionary-conserved allosteric properties of three neuronal calcium sensor proteins. *Front Mol Neurosci*. 2019;12:50.
- Marino V, Sulmann S, Koch KW, Dell'Orco D. Structural effects of Mg(2)(+) on the regulatory states of three neuronal calcium sensors operating in vertebrate phototransduction. *Biochim Biophys Acta*. 2015;1853:2055–65.
- Montemayor C, Montemayor OA, Ridgeway A, Lin F, Wheeler DA, Pletcher SD, et al. Genome-wide analysis of binding sites and direct target genes of the orphan nuclear receptor NR2F1/COUP-TFI. *PLoS One*. 2010;5:e8910.
- Neuman K, Soosaar A, Nornes HO, Neuman T. Orphan receptor COUP-TF I antagonizes retinoic acid-induced neuronal differentiation. *J Neurosci Res*. 1995;41:39–48.
- Nooren IM, Thornton JM. Diversity of protein-protein interactions. *EMBO J*. 2003;22:3486–92.
- Park JI, Tsai SY, Tsai MJ. Molecular mechanism of chicken ovalbumin upstream promoter-transcription factor (COUP-TF) actions. *Keio J Med*. 2003;52:174–81.
- Pereira FA, Tsai MJ, Tsai SY. COUP-TF orphan nuclear receptors in development and differentiation. *Cell Mol Life Sci*. 2000;57:1388–98.
- Perlmann T, Umehono K, Rangarajan PN, Forman BM, Evans RM. Two distinct dimerization interfaces differentially modulate target gene specificity of nuclear hormone receptors. *Mol Endocrinol*. 1996;10:958–66.
- Pinaire J, Hasanadka R, Fang M, Chou WY, Stewart MJ, Kruijjer W, et al. The retinoid X receptor response element in the human aldehyde dehydrogenase 2 promoter is antagonized by the chicken ovalbumin upstream promoter family of orphan receptors. *Arch Biochem Biophys*. 2000;380:192–200.
- Quintero CM, Laursen KB, Mongan NP, Luo M, Gudas LJ. CARM1 (PRMT4) acts as a transcriptional coactivator during retinoic acid-induced embryonic stem cell differentiation. *J Mol Biol*. 2018;430:4168–82.
- Rastinejad F, Ollendorff V, Polikarpov I. Nuclear receptor full-length architectures: confronting myth and illusion with high resolution. *Trends Biochem Sci*. 2015;40:16–24.
- Rech ME, McCarthy JM, Chen CA, Edmond JC, Shah VS, Bosch DGM, et al. Phenotypic expansion of Bosch-Boonstra-Schaaf optic atrophy syndrome and further evidence for genotype-phenotype correlations. *Am J Med Genet A*. 2020;182:1426–37.
- Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ, et al. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature*. 2012;485:237–41.
- Schaaf C, Yu-Wai-Man P, Valentin I. NR2F1-related neurodevelopmental disorder. In: Adam MP, Feldman J, Mirzaa GM, Pagon RA, Wallace SE, Bean LJH, et al., editors. *GeneReviews*® [Internet]. Seattle (WA): University of Washington; 1993.
- Schneider CA, Rasband WS, Eliceiri KW. NIH image to ImageJ: 25 years of image analysis. *Nat Methods*. 2012;9:671–5.
- Seidel L, Zarzycka B, Zaidi SA, Katritch V, Coin I. Structural insight into the activation of a class B G-protein-coupled receptor by peptide hormones in live human cells. *eLife*. 2017;6:e27711.
- Sessler RJ, Noy N. A ligand-activated nuclear localization signal in cellular retinoic acid binding protein-II. *Mol Cell*. 2005;18:343–53.
- Tang K, Tsai SY, Tsai MJ. COUP-TFs and eye development. *Biochim Biophys Acta*. 2015;1849:201–9.
- Tocco C, Bertacchi M, Studer M. Structural and functional aspects of the neurodevelopmental gene NR2F1: from animal models to human pathology. *Front Mol Neurosci*. 2021;14:767965.
- Tran P, Zhang XK, Salbert G, Hermann T, Lehmann JM, Pfahl M. COUP orphan receptors are negative regulators of retinoic acid response pathways. *Mol Cell Biol*. 1992;12:4666–76.
- Vaezeslami S, Mathes E, Vasileiou C, Borhan B, Geiger JH. The structure of apo-wild-type cellular retinoic acid binding protein II at 1.4 Å and its relationship to ligand binding and nuclear translocation. *J Mol Biol*. 2006;363:687–701.
- Walsh S, Gosswein SS, Rump A, von der Hagen M, Hackmann K, Schrock E, et al. Novel dominant-negative NR2F1 frameshift mutation and a phenotypic expansion of the Bosch-Boonstra-Schaaf optic atrophy syndrome. *Eur J Med Genet*. 2020;63:104019.
- Zhang Y, Dufau ML. EAR2 and EAR3/COUP-TFI regulate transcription of the rat LH receptor. *Mol Endocrinol*. 2001;15:1891–905.
- Zhuang Y, Gudas LJ. Overexpression of COUP-TF1 in murine embryonic stem cells reduces retinoic acid-associated growth arrest and increases extraembryonic endoderm gene expression. *Differentiation*. 2008;76:760–71.



Zou W, Cheng L, Lu S, Wu Z. A de novo nonsense mutation in the N-terminal of ligand-binding domain of NR2F1 gene provoked a milder phenotype of BBSOAS. *Ophthalmic Genet.* 2020;41:88–9.

### SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Marino V, Phromkrasae W, Bertacchi M, Cassini P, Chakrabandhu K, Dell'Orco D, et al. Disrupted protein interaction dynamics in a genetic neurodevelopmental disorder revealed by structural bioinformatics and genetic code expansion. *Protein Science.* 2024;33(4):e4953. <https://doi.org/10.1002/pro.4953>