Landscape of mSWI/SNF chromatin remodeling complex perturbations in neurodevelopmental disorders

Alfredo M. Valencia[1,2,3,10,11,12], Akshay Sankar[1,3,12], Pleuntje J. van der Sluijs [4], F. Kyle Satterstrom [5,6], Jack Fu[6], Michael E. Talkowski [5,6], Samantha A. Schrier Vergano [7,8], Gijs W. E. Santen [4] & Cigall Kadoch [1,3,9] ✉

DNA sequencing-based studies of neurodevelopmental disorders (NDDs) have identified a wide range of genetic determinants. However, a comprehensive analysis of these data, in aggregate, has not to date been performed. Here, we find that genes encoding the mammalian SWI/SNF (mSWI/SNF or BAF) family of ATP-dependent chromatin remodeling protein complexes harbor the greatest number of de novo missense and protein-truncating variants among nuclear protein complexes. Non-truncating NDD-associated protein variants predominantly disrupt the cBAF subcomplex and cluster in four key structural regions associated with high disease severity, including mSWI/SNF-nucleosome interfaces, the ATPase-core ARID-armadillo repeat (ARM) module insertion site, the Arp module and DNA-binding domains. Although over 70% of the residues perturbed in NDDs overlap with those mutated in cancer, ~60% of amino acid changes are NDD-specific. These findings provide a foundation to functionally group variants and link complex aberrancies to phenotypic severity, serving as a resource for the chromatin, clinical genetics and neurodevelopment communities.

Sequencing studies have revealed extensive involvement of chromatin regulatory processes in a range of human diseases, with frequent mutations in the genes encoding proteins that govern chromatin architecture[1–4]. Four families of multi-subunit ATP-dependent chromatin remodeling complexes (SWI/SNF, ISWI, CHD and INO80) modulate chromatin topology and gene expression by mobilizing their nucleosome substrates[5]. Recent advances in cryo-electron microscopy (cryo-EM), cross-linking mass spectrometry and homology modeling have begun to uncover the three-dimensional (3D) structure and modes of nucleosome substrate engagement of these large heterogeneous entities, informing mechanistic studies[6].

Mutations in the genes encoding mammalian SWI/SNF (mSWI/SNF) chromatin remodeling complex are found in over 20% of cases in cancer, which has stimulated a range of basic and translational efforts over the past several years[7–9]. A wealth of mutational data of neurodevelopmental disorders (NDDs), such as intellectual disability and autism

[1]Dana-Farber Cancer Institute and Harvard Medical School, Boston, MA, USA. [2]Chemical Biology Program, Harvard University, Cambridge, MA, USA. [3]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [4]Department of Clinical Genetics, Leiden University Medical Center, Leiden, the Netherlands. [5]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [6]Massachusetts General Hospital, Boston, MA, USA. [7]Children's Hospital of the King's Daughters, Norfolk, Virginia, USA. [8]Department of Pediatrics, Eastern Virginia Medical School, Norfolk, Virginia, USA. [9]Howard Hughes Medical Institute, Chevy Chase, MD, USA. [10]Present address: Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA, USA. [11]Present address: Stanford Brain Organogenesis, Wu Tsai Neurosciences Institute, Stanford University, Stanford, CA, USA. [12]These authors contributed equally: Alfredo M. Valencia, Akshay Sankar. ✉e-mail: Cigall_kadoch@dfci.harvard.edu

spectrum disorders, has also recently emphasized a high mutational burden of chromatin regulatory genes in NDD, presenting an opportunity to dissect the molecular underpinnings and to inform potential strategies to remedy the comorbid issues associated with these disorders[2,10–14].

Most cancer-associated mSWI/SNF mutations result in subunit deletions or gene silencing, which has presented the field with opportunities to understand the impact of full subunit losses and the impact on complex disassembly[15–18]. NDD-associated mSWI/SNF genetic variants present particularly unique opportunities for functional dissection, in that 1) mutations are often missense, affecting single amino acids and clustering in defined domains within subunits; 2) mutations are predominantly heterozygous, underscoring the high degree of dosage sensitivity; and 3) mutations are often found as the sole genetic cause of these disorders. Furthermore, for trios in which parents' genetic information is available, mSWI/SNF gene variants are predominantly de novo (absent in parents), indicating their causative role[19–21]. Together, these features enable functional assignment and prioritization for specific subunit domains and even individual protein residues. Identifying and mechanistically defining these variants will be critical for the assignment of specific chromatin remodeling complex functions and, ultimately, informing therapeutic approaches for a range of human diseases driven by mSWI/SNF complex disruption.

Here, we sought to comprehensively catalog and integrate mSWI/SNF complex sequence variants across a diverse collection of datasets, including the Simon's Foundation Research Initiative (SFARI) (Simons Foundation Powering Autism Research for Knowledge (SPARK), Simons Searchlight Collection–Autism Sequencing Consortium (SSC-ASC)), the Deciphering Developmental Disorders project (DDD), the DECIPHER database[22], ClinVar[23], the Leiden Open Variation Database (LOVD)[24], de novo sequence variants from the literature (as performed in McRae et al. (https://github.com/jeremymcrae/dnm_cohorts)[3,25–39]), NDD-associated mSWI/SNF sequence variants from the literature[3,19–21,35,40–81] and 85 previously unreported NDD-associated mSWI/SNF cases, including 72 novel variants, focused on protein coding mutations stemming from single-nucleotide variants (SNVs) and small insertions/deletions (indels) (Supplementary Table 1). These analyses encompass 2,539 total cases of which the majority (67.1%, $n = 1,703$) result in missense and in-frame indels that collectively reveal 1,204 unique variants.

Results

Chromatin remodelers carry a high mutational burden in NDDs

Single amino acid mutations and protein-truncating variants (PTVs) in chromatin regulatory genes are pathogenic for a variety of NDDs, including syndromic and non-syndromic intellectual disabilities and autism spectrum disorders[3], but their relative prevalence remains undefined. We collated and analyzed all SNVs and small indels reported in DECIPHER (DatabasE of genomiC varIation and Phenotype in Humans using Ensembl Resources)[22] (https://www.deciphergenomics.org/), a repository of clinical and genetic information on individuals with developmental disorders. Remarkably, we found that epigenetic and chromatin-related genes (EpiFactor gene list, Supplementary Table 2)[82] were more frequently mutated than synapse-related genes (SynGO gene list, Supplementary Table 2)[83], which are known to be highly implicated in NDDs (Extended Data Fig. 1a). By examining the top 50 Gene Ontology molecular functions (GOMFs) of genes in the Development Disorder Genotype–Phenotype Database (DDG2P), we found that top-ranked disrupted processes were enriched for transcription- and chromatin-related processes, with transcription and chromatin binding terms ranking highest among them (Fig. 1a and Extended Data Fig. 1b,c). Performing this analysis with variants identified from the SFARI Autism Spectrum Disorder (ASD) SPARK, SSC-ASC and developmental disorder (DD) DDD study datasets (ASD + DD) revealed similar results, including transcription-, synapse- and chromatin-related GOMFs (that is, 1: transcriptional coregulator activity, 2: voltage-gated channel activity,

3: voltage-gated cation channel activity and 4: chromatin DNA binding) (Extended Data Fig. 1a,c–e). We then analyzed de novo missense and PTV frequencies from ASD + DD datasets by protein complex associations and by chromatin regulatory activity, which revealed the greatest number of variants occurred in SWI/SNF chromatin remodeling complex genes (protein complex, $n = 404$ sequence variants, rank 1), followed by SET1 methyltransferase family (protein family, $n = 346$, rank 2), lysine acetyltransferases (protein family, $n = 300$, rank 3) and CHD chromatin remodeling complex genes (protein complex, $n = 232$, rank 4) (Fig. 1b and Supplementary Table 2). This result was consistent using DECIPHER data (Extended Data Fig. 1f) and chromatin-related protein complexes from EpiFactor using ASD + DD data (Extended Data Fig. 1g). Of note, several histone modifying complexes, including the histone–lysine N-methyltransferase (KMT2 or MLL) family of complexes, the histone acetyltransferase MOZ/MORF complexes and Polycomb repressive deubiquitinase (PR-DUB) complexes had a greater average of mutations when normalized by gene set size, owing to lower numbers of defined components relative to mSWI/SNF complexes (average ~6 components versus ~19 components for mSWI/SNF) (Extended Data Fig. 1h and Supplementary Table 2). Nevertheless, when normalized by protein length (or gene exon length), cBAF complexes maintained the highest average number of de novo mutations and PTVs compared to all EpiFactor complexes (Extended Data Fig. 1i). Interestingly, separating ASD and DD datasets revealed cBAF was the most frequently mutated gene set in DD but ranked fourth in ASD, potentially suggesting a subtle distinction between ASD-associated variants from SFARI compared to a mixture of ASD and other NDDs reported in the DDD database (Extended Data Fig. 1j).

Expanding our analysis to include copy-number variants in addition to SNVs/indels using DECIPHER, we found that genes encoding all members of mammalian chromatin remodeling complexes (across all families) are implicated in approximately one in ten of all DECIPHER cases (9.34%, 5,196/55,645) (Fig. 1c,d and Extended Data Fig. 1k). The 29 genes encoding the mSWI/SNF complex are affected in the greatest percentage (4.10%, 2,281/55,645), the majority of which are classified as 'pathogenic' or 'likely pathogenic' (67.9%, 1,548/2,281), 39.2% of which were confirmed de novo and 34.4 % of unknown inheritance (Extended Data Fig. 1l). Many mSWI/SNF genes are also implicated in ASD, as characterized by SFARI database (Fig. 1d)[84]. Notably, genes such as ARID1B, SMARCA4 and SMARCA2 were among the top mSWI/SNF genes with most de novo missense and PTVs across all ASD + DD cases, with ARID1B having the most variants, followed by ANKRD11, KMT2A, and SCN2A (Extended Data Fig. 1m–n). When including CNV losses and sequence variants from DECIPHER, the top mSWI/SNF genes implicated were SMARCB1 and SMARCA2, mutations in which cause the most severe phenotypes of mSWI/SNF-related NDDs, CSS and Nicolaides-Baraitser syndrome (NCBRS), respectively[85] (Fig. 1c). Nevertheless, multiple genes may be disrupted in a given CNV, making genotype-phenotype correlations more challenging to directly assess. As compared to cancer, wherein mutations in mSWI/SNF genes are present in 20.3% of all cases sequenced[86] (COSMIC: the Catalog of Somatic Mutations in Cancer), specific mSWI/SNF subunits were more frequently mutated in NDD relative to other mSWI/SNF genes. These included ARID1B, the paralog of which, ARID1A, is among one of the most frequently mutated genes in cancer, SMARCA4, and SMARCA2 (Extended Data Fig. 1o). Notably, genes encoding PBAF and ncBAF components such as PBRM1, ARID2, BICRAL (GLTSCR1L) and others were found to be more frequently mutated in cancer than in NDD (Extended Data Fig. 1p). As the most frequently mutated chromatin remodeler in NDDs and cancer, the remainder of this Analysis is centered on the mSWI/SNF family of chromatin remodeling complexes.

mSWI/SNF NDD variants accumulate in functional domains

To comprehensively examine the full constellation of mSWI/SNF sequence variants in NDD, we combined mSWI/SNF gene mutations

Fig. 1 | Genes encoding chromatin regulatory complexes represent the most frequently mutated gene classes in human NDDs. a, Scatterplot of the average numbers of non-benign SNVs in DECIPHER corresponding to the top 50 GOMF gene sets enriched in DDG2P developmental disorder-associated genes, ranked by the mutational burden of each gene set. **b**, Bar graph depicting the total number of NDD-associated missense and protein truncating variants (PTVs) for a curated list of chromatin regulatory and epigenetic gene sets, ranked by mutational burden of each gene set in autism spectrum disorders and developmental disorders (ASD + DD) from the Simons Foundation Research Initiative (SFARI) datasets (SPARK: Simons Foundation Powering Autism Research + SSC-ASC: Simons Searchlight Collection–Autism Sequencing Consortium, and DDD: Deciphering developmental disorders studies). The mSWI/SNF chromatin remodeling complex gene set is emphasized in red. **c**, Heatmaps depicting the mutational frequency for genes encoding members of the SWI/SNF, CHD, ISWI, and INO80 complex families in DECIPHER. Total number

of variants (including copy-number variant (CNV) gain, copy number variant loss (CNV loss), single nucleotide variant (SNV) and indel mutational frequencies are indicated. Percentage of total DECIPHER sequence variants are indicated for each chromatin remodeling complex family (top). **d**, Cartoon representations of the four classes of chromatin remodelers (SWI/SNF, CHD, ISWI and INO80) and respective subcomplex or related complex associations, colored by CNV loss/ SNV/indel variation frequency from panel **c**. Interchangeable subunit paralogs are colored by their combined mutational frequency. Autism spectrum disorder (ASD) risk score (SFARI) and developmental disorder associations curated from literature and OMIM (Online Mendelian Inheritance in Man, a catalog of human genes and genetic disorders; https://www.omim.org/) are indicated. Asterisk (*) indicates paralog implicated in NDD. Where possible, cartoons were based on 3D structural data available from human and yeast structures; ovals are used in in lieu of structural cartoons for components lacking structural data.

from the DECIPHER, ClinVar, LOVD, SFARI SPARK and SSC–ASC datasets and merged these with mutations reported in published literature as well as n = 85 novel, previously unreported cases (Supplementary Table 1). After removing duplicates, variants with a mutant allele frequency of >0.5% in the general population as assessed by gnomAD[87], and filtering for missense, inframeshift (herein defined as non-frameshift inducing insertions/deletions), frameshift and nonsense variants, we identified 2539 variants in mSWI/SNF genes, 61.5% of which were missense

(Fig. 2a). Variants resulted predominantly in missense or inframeshift (67.1%) (Fig. 2a,b), with the exception of *ARID1B* and *ARID2*, for which the majority of variants were nonsense or frameshift (Fig. 2b). The greatest number of missense variants stemmed from G > A and C > T base pair conversions, resulting in a variety of amino acid changes (Extended Data Fig. 2a–e). The most frequently altered residues were Arginine (R), Proline (P), Alanine (A), and Glycine (G), together making up 47% (815/1703) of all missense and inframeshift affected residues

in the dataset (Extended Data Fig. 2b–e, Supplementary Table 1). Furthermore, the most common missense amino acid substitution was Arginine to Histidine (Arg>His; R > H), indicating reductions in both the relative size and pKa of the amino acid side chain (Arg pKa 12.48 – His pKa 6.0) (Extended Data Fig. 2e).

A high percentage of missense and indel mSWI/SNF mutations localized to highly conserved regions (53.1% high, 24.7% moderate conservation) (Fig. 2c). Mutations in subunits such as ACTB, ACTL6A/B, DPF2, and SMARCB1 entirely or nearly entirely occurred in intra-domain structured regions, whereas variants in BCL7A/B, PHF10, and ARID1A/B subunits were skewed toward interdomain disordered regions (Fig. 2d, Extended Data Fig. 2f and Supplementary Table 3). Intriguingly, mutations in SMARCA2 clustered in the ATPase/helicase domain, whereas mutations in SMARCA4 were more dispersed throughout the protein, including the structurally unresolved N terminus (Fig. 2e). Interestingly, whereas mutations within the SMARCA2 helicase cause NCBRS, SMARCA2 mutations outside of this domain are implicated in a distinct disorder, blepharophimosis-impaired intellectual disability syndrome[88]. Among mSWI/SNF paralogs, frameshift mutations were more enriched in ARID1B, whereas missense mutations in specific regions were enriched in ARID1A, clustering namely in the ARID DNA-binding domain, the structurally unresolved N terminus and the C-terminal armadillo repeat domain (ARM or core binding region) (Fig. 2e). A possibility underlying this difference is that ARID1A haplo-insufficient mutations lead to a more severe phenotype, as suggested by the frequent occurrence of mosaic variants[69] and further substantiated during the review process by an analysis of fetal cases[89].

Genotype-phenotype clinical studies have suggested that *ARID1B* truncating mutations are generally linked to the mildest cases of CSS-related intellectual disability, including some individuals without intellectual disability[90], whereas single amino acid mutations of the SMARCB1 protein are correlated with the most severe cognitive impairment and growth delay in CSS[21,69,85]. SMARCA2-ATPase mutations result in severe intellectual disability cases of NCBRS, but SMARCE1-HMG and DPF2-PHD mutations are correlated to moderate-severe and mild intellectual disability phenotypes, respectively[72,74,91]. We examined non-truncating variants through predicted phenotypic severity score analysis (PolyPhen HumVar[92]), which highlighted domains such as the SMARCB1-CTD, ARID2-ARID and SMARCA2-Helicase-C and SMARCA2-post-Helicase-C as those predicted to result in most severe disease phenotypes, in agreement with published phenotypic data (Fig. 2f and Supplementary Table 3). This analysis also highlighted the SMARCC1-post-SWIRM interdomain with a particularly high PolyPhen score and average number of mutations; this region lacks 3D structural definition, implicating an alternative contribution to mSWI/SNF function (Fig. 2f). Collectively, these results highlight convergent clinical outcomes stemming from mSWI/SNF gene disruption, with variation in severity observed across distinct proteins and even domains of mSWI/SNF complex components.

Mapping NDD missense/inframeshift variants on 3D SWI/SNF-nucleosome models

We next integrated these sequence variant data with recently solved structures of mSWI/SNF cBAF complexes[93,94], which allowed for mapping of 238 unique positions comprising 44.08% (655/1,486) of the theoretically mappable cBAF-specific NDD missense and in-frame indels on the recombinant cBAF cryo-EM structure, and 51.55% (766/1,486) on the endogenous structure for all cBAF paralogs (Fig. 3, Extended Data Fig. 3a, b and Supplementary Table 3)[95,96]. These results highlight the need for further structural efforts as well as studies to define the roles and interactions of non-structured, disordered regions. Mapping subcomplex-specific positions onto the recently solved PBAF complex bound to a nucleosome[97] resolved 20 additional PBAF-specific subunit mutations across ARID2, PBRM1 and BRD7 (Extended Data Fig. 3c). For ARID1B, SMARCA2 and ACTL6B, paralog subunits that are not part of the solved protein complex, we mapped mutant residues on to the respective paralogs following paralog alignment (Fig. 3 and Extended Data Fig. 3a).

This structural analysis reveals that BAF complex compromises in NDD cluster primarily in four distinct regions on mSWI/SNF complexes: the catalytic ATPase module, the mSWI/SNF core, the Arp module, and the SMARCB1 BAF-nucleosome contact point (Fig. 4a–d). As demonstrated initially through our previous work[98] and later resolved in 3D structural efforts, CSS-associated mutations in SMARCB1 localize to the SMARCB1-CTD, the key and only interface connecting the mSWI/SNF core module to the nucleosome acidic patch (Fig. 4a and Extended Data Fig. 4a). Second, mutations in the SMARCA4 ATPase subunit are primarily situated in the ATP-coordinating and DNA-binding residues near the nucleosome, with additional mutations accumulating within the region of SMARCA4 interfacing within the mSWI/SNF core (Fig. 4b,c and Extended Data Fig. 4b). We also identified a cluster of variants are found throughout the ACTB subunit of the Arp module, whose mutation is associated with severe cases of Baraitser-Winter cerebrofrontofacial syndrome[75,95] (Fig. 4d).
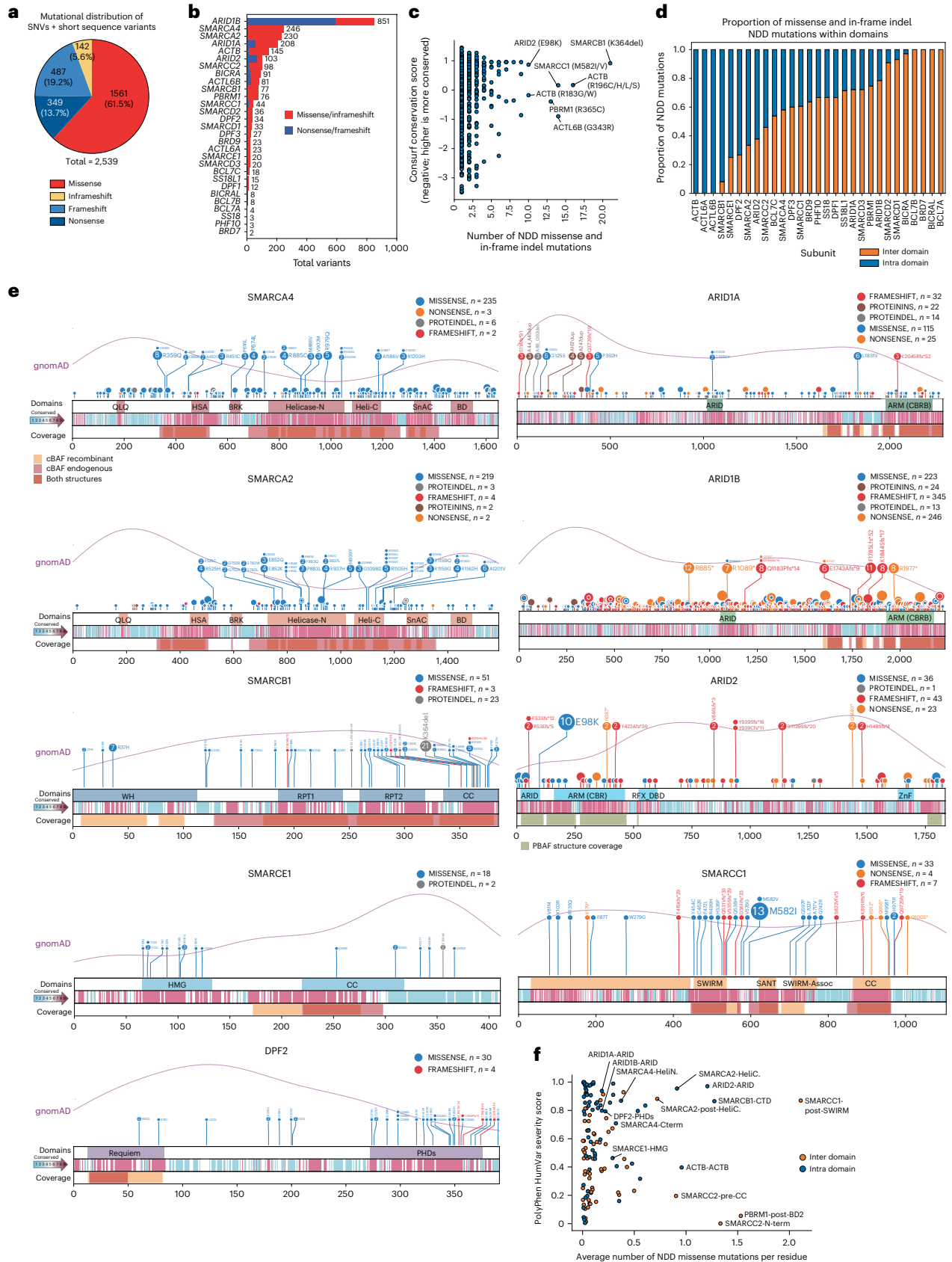
Intriguingly, whereas mutations to positively charged residues within the SMARCB1-CTD disrupt binding to the nucleosome and result in severe intellectual disability[93,94,98], we report two novel variants in the SMARCB1-CTD, D369E and R376K, in which a positive or negative charge is maintained, and which are phenotypically associated with less severe disease (Fig. 4a, red, and Supplementary Table 1), underscoring that defining chemical properties of distinct mutations, even within a given subunit domain, may inform intellectual disability severity and phenotypic outcomes.

We next mapped cBAF NDD-mutant residues by amino acid characteristics (that is, charged, polar, nonpolar, etc). This map highlighted that many NDD-associated ACTB residues are nonpolar, the mutation of which is predicted to disrupt hydrophobic core as further suggested by Missense3D[96,99] (Extended Data Fig. 4c and Supplementary Table 3; http://missense3d.bc.ic.ac.uk/). Within the context of mSWI/SNF (ACTB

Fig. 2 | Analysis of NDD-associated SNV and indel mutations in mSWI/SNF complex components. a, Pie chart reflecting the distribution of n = 2,539 mSWI/SNF NDD-associated SNV and in-frame indel mutations from an integrated dataset containing data from SPARK, SSC-ASC, DDD, DECIPHER, ClinVar, LOVD, literature curation and novel variants reported in this study. **b**, Bar chart summarizing total NDD-associated missense/in-frame deletions and insertions (red) and nonsense/frameshift-inducing mutations (blue) across all mSWI/SNF genes. **c**, Scatterplot of the negative-normalized ConSurf conservation score versus the mutational recurrence at each mSWI/SNF complex subunit residue for NDD missense and in-frame variants in the integrated dataset. Highly conserved and highly mutated positions are labeled. **d**, Stacked bar chart summarizing proportion of NDD-associated missense and in-frame insertion/deletion variants falling within (intra, blue) or outside (inter, orange) of mSWI/SNF subunit domains in the integrated dataset. Domains annotated by PFAM, UniProtKB, manual curation, and structurally resolved domains (see also Supplementary Table 3). **e**, Lollipop plots

of NDD mutations in the integrated dataset across protein domain schematics of ARID1A/B, ARID2, SMARCA2/4, SMARCB1, SMARCC1, SMARCE1, and DPF2 subunits generated with Protein Paint. Missense (blue), nonsense (orange), frameshift (red), in-frame deletions (gray) and insertions (brown) are shown. Kernel density estimates (relative frequency distribution) of gnomAD missense mutations (purple line) are overlaid. Domain annotations informed by PFAM, UniProtKB, manual curation, or by structurally resolved domains are indicated. ConSurf conservation scores are shown in a cyan-white-magenta heatmap in increasing conservation order, and structural coverages of the nucleosome core particle (NCP)-bound human cBAF (light orange, PDB: 6LTJ), endogenous human cBAF-NCP bound (red, PDBDEV00000056), and by both structures (brown). Structural coverage for the NCP-bound PBAF complex is also shown for ARID2 (light green, PDB:7VDV). **f**, PolyPhen HumVar predicted phenotypic severity score and missense mutational recurrence of mSWI/SNF gene mutations from the integrated dataset in intra (blue) and inter (orange) domains.

is also a member of INO80 and TIP60 complexes; Supplementary Table 2), ACTB mutations are predicted to alter buried hydrophobic cavities, as well as interaction with the ACTL6A Arp module binding partner, and even the HSA helix of the SMARCA4 ATPase (Extended

Data Fig. 4d). Intriguingly, some of the most recurrent ACTL6A and ACTL6B mutations of the Arp module, R377W and G343R, are located in close proximity to one another when mapped onto ACTL6A subunit on the cBAF structure (Fig. 4d). Although not interfacing other

SMARCA4 (ATPase)
T1170A
S1176G
A1186V/T
A1190V
V1199M
V1201M
R1203H/C
R1308W
H1309R
I1347N
D1351N
G784R
S813L
A826S
R841I
R842Q/W
T859M
K865E
R874H
G883S
R885C/H
M886V
C891Y
T894M
V902M
R906C
P913L
L921F
C936S
R978Q
R979Q

ACTL6A
I18T
I70T
R76C
M92T/I
H105R
M122L
P126L
T175I
I340V
R377W
R389W

SMARCA4
N477D
R359Q/W
I366M/F
R370C
L430P
N435D
E449K
A450S/D
R451C/P
T453I
E457D
K458_E465del

SMARCA4 (ATPase)
R1043Q/W/L
H1010R
M1011T
S1079L
L1092R
L1108P
M1109V
M1112T
A1117V
R1119C
G1120S
T1156A
R1157Q/G
D1169G

SMARCD1
Q145P
P293T
D330E
I428V
R446G
V470del
A487G
F495L

SMARCE1
K256R

DPF2
R6OQ
R78C

SMARCB1
G11R
R37H
G271E
R341W
T357R
K363N/E
K364del
R366C/G
D369V/E
R374Q/W
R377H
I28M
E184K
D196N
R201Q
E216K
A238T
V262I
H269R
A294T
K296N
L301P
E362_I365delinsV
D367G
R376K

ARID1A
T1649R
P1651L
R1656K
D1697G
G1711E
L1713P
L1731S
G1737S
T1743M
N1820D
L1831V
R1833H
I1988T
F1993_V1994dup
N1997D
G2012R
L2061S
N2066Y
E2078K
V2084G
H2090R
A2137D
R2164W
L2171dup
N2173I
G2177R
M2231T
R2233Q
R2236C
P2265L
N2194_L2195delinsGN

SMARCC2
R443Q
A445T
L609P
L610P
L613P
C635R
A868T
E893G
M896V
E900G
E921K

ACTB
D2Y
D3F
A7T/V
D11del
N12H/D
A22T
D24N
A26V
R28G
F31L
H87D
F90del
N92S
V103L
A108D
P109S
E117V/D/K
M119T
T120I
I136V
L140V
S145C
T149I
P164S/H
L171F
H173Y
L176P
D179E
R183W/G
D184Y
T194A
R196H/C/L/S
G197S
A204G
R206Q/W
V209M
R210H/C
K215N
T229M
P243L
G245S
P258L
A260E
G268R/C
Y294D
N296K
V298A
G302A
M305T
K315del
P322R
I330L
R335P
S338F
I341V
S348L
W356R
E364K
H371R

BAF NDD mutations
- 🔴 Primary subunit (in structure)
- 🔵 Paralog mapped
- 🟣 In both primary and paralog

Red text: indicates recurrence of 3 or greater
Bold text: indicates newly reported case

Fig. 3 | Mapping of 238 unique NDD-associated variant positions onto the structure of the human cBAF complex. NDD-associated variants, including 14 novel variants, mapped on to the 3D structure of the human cBAF complex (PDB:6LTJ). Residues shown in red spheres represent NDD-associated variants in the subunit indicated, residues in blue represent those mapped from the paralog subunit, and residues in purple represent NDD-variants mapped in both the primary subunit present on the cBAF structure and paralog mapped subunit. Recurrent variants ($n \geq 3$) are emphasized in red text. Caution is needed when evaluating these variants in a clinical context since not all variants are confirmed as causal.

mSWI/SNF subunits, these residues are oriented toward the DNA exit, and we hypothesize that the ACTL6A-R377 residue may stably bind the DNA backbone adjacent to the nucleosome, which would be disrupted upon mutation to a nonpolar residue such as tryptophan (R377W). Conversely, the addition of a positive charge in ACTL6B from side chain-absent glycine (G) to arginine (R) upon mutation may impart affinity to the nucleosomal DNA.

We predicted that SMARCB1 mutations in the RPT2 domain may disrupt the RPT domain cavity (Extended Data Fig. 4e). Further, the recurrent SMARCB1-R37H mutation in the winged-helix DNA-binding domain, which causes severe intellectual disability and Kleefstra-like syndrome, also demonstrated hydrogen bonding with the carbonyl backbone of ARID1A-L2073 and Y2076 that is likely disrupted upon mutation (Extended Data Fig. 4e). Intriguingly, the SMARCB1-WH

domain is isolated from the SMARCB1 C-terminus on the recombinant cBAF structure but is predicted to be repositioned closer to the nucleosome binding lobe in the PBAF structure[97], suggesting potentially distinct roles and functional impacts of the SMARCB1-R37H mutation in cBAF compared to PBAF, perhaps independent of remodeling activity as the SMARCB1-R37H mutation does not impact cBAF nucleosome remodeling activity in vitro[98].

Yeast SWI/SNF ATPases offer NDD variant functional insights

Given the high frequency of mutations within the catalytic ATPase subunits of mSWI/SNF chromatin remodeling complexes, SMARCA2 and SMARCA4, we mapped conserved mutant residues onto the nucleosome-bound yeast SWI/SNF and SNF2 structures[100,101] (Extended Data Fig. 4f). Interestingly, the current human cBAF structures do not
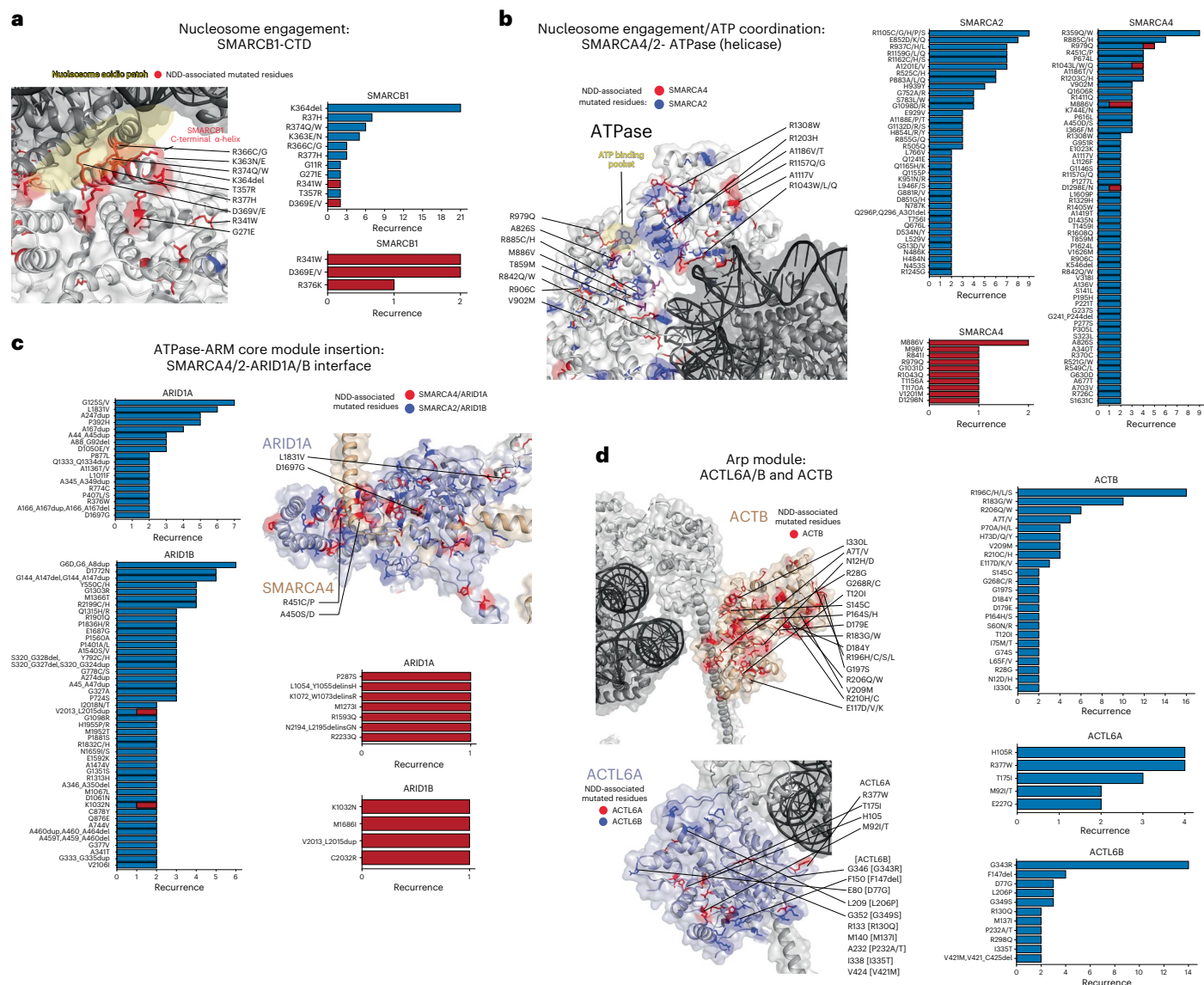
Fig. 4 | NDD-associated mutations cluster within key structural hubs of mSWI/SNF complexes. a, Zoomed-in view of the SMARCB1 C-terminal alpha-helix domain (PDB:6LTJ) with the nucleosome acidic patch interaction site highlighted in yellow (left). NDD-associated mutations in SMARCB1 are emphasized in red. All NDD-associated SMARCB1–C terminal alpha-helix mutations ranked by frequency (right). Novel SMARCB1 variant cases reported in this study shown in red bar chart. **b**, Zoomed-in view of the SMARCA4 ATPase subunit within the cBAF complex (PDB:6LTJ) at its interface with the nucleosome (left). Mutations in SMARCA4 are indicated in red; mutations in SMARCA2 are indicated in blue, shared mapped in purple. ATP binding pocket is highlighted in yellow. NDD-associated missense and inframeshift variants in SMARCA4 and SMARCA2, ranked by frequency, filtered for recurrence of $n \geq 2$ by position (right). Novel SMARCA4 cases reported in this study shown in red bar chart.

c, NDD-associated mutations in ARID1A and ARID1B, ranked by frequency, filtered for recurrence of $n \geq 2$ by position (left). Zoomed-in view of the SMARCA4-ARID1A interface within the core module of the cBAF complex (right). SMARCA4 is shown in tan and ARID1A in light purple, with mutations in SMARCA4 and ARID1A shown in red and those in their respective paralogs SMARCA2 and ARID1B shown in blue. Novel ARID1A/B variant cases reported in this study shown in red bar chart. **d**, Left, zoomed-in view of the ACTB (tan) and ACTL6A (light purple) subunits within the Arp module of the cBAF complex, with mutations indicated in red and blue for ACTL6A paralog subunit, ACTL6B. NDD-associated mutations in ACTL6A, ACTL6B and Actin, ranked by frequency, filtered for recurrence of $n \geq 2$ by position (right). Recurrent ACTL6B variants donated in brackets mapped onto ACTL6A indicated.

resolve the brace helices, and we highlight residues that are buried in the brace helices (SMARCA4 978-979) (Extended Data Fig. 4f). Cancer- and NDD-associated mutations (R973W and R1243W) in the brace helices of SMARCA4 were recently found to diminish nucleosome remodeling activity of PBAF complexes in vitro[97]. Given their proximity to this region and the ATP pocket of SMARCA4, we posit that additional variants in the brace helices and the nearby R978Q and R979Q variants would have similar deficits in nucleosome remodeling in human cells (Extended Data Fig. 4g). To assess the potential impact

that NDD-associated mutations might have on ATP engagement, given that structures are static, we mapped conserved SMARCA2/4 mutant residues onto the open state, ADP bound (similar to apo structure) and onto the closed, ADP-BeFx-bound yeast SNF2 nucleosome bound structures[102], which allows mapping of ~85% of all SMARCA2/4-ATPase positions (Extended Data Fig. 4h). Furthermore, this mapping highlighted NDD-associated nucleosome binding residues such as N1050 and K1057 (corresponding NDD variants: SMARCA2-N1007K and K1044E), which were previously shown to dramatically diminish nucleosome
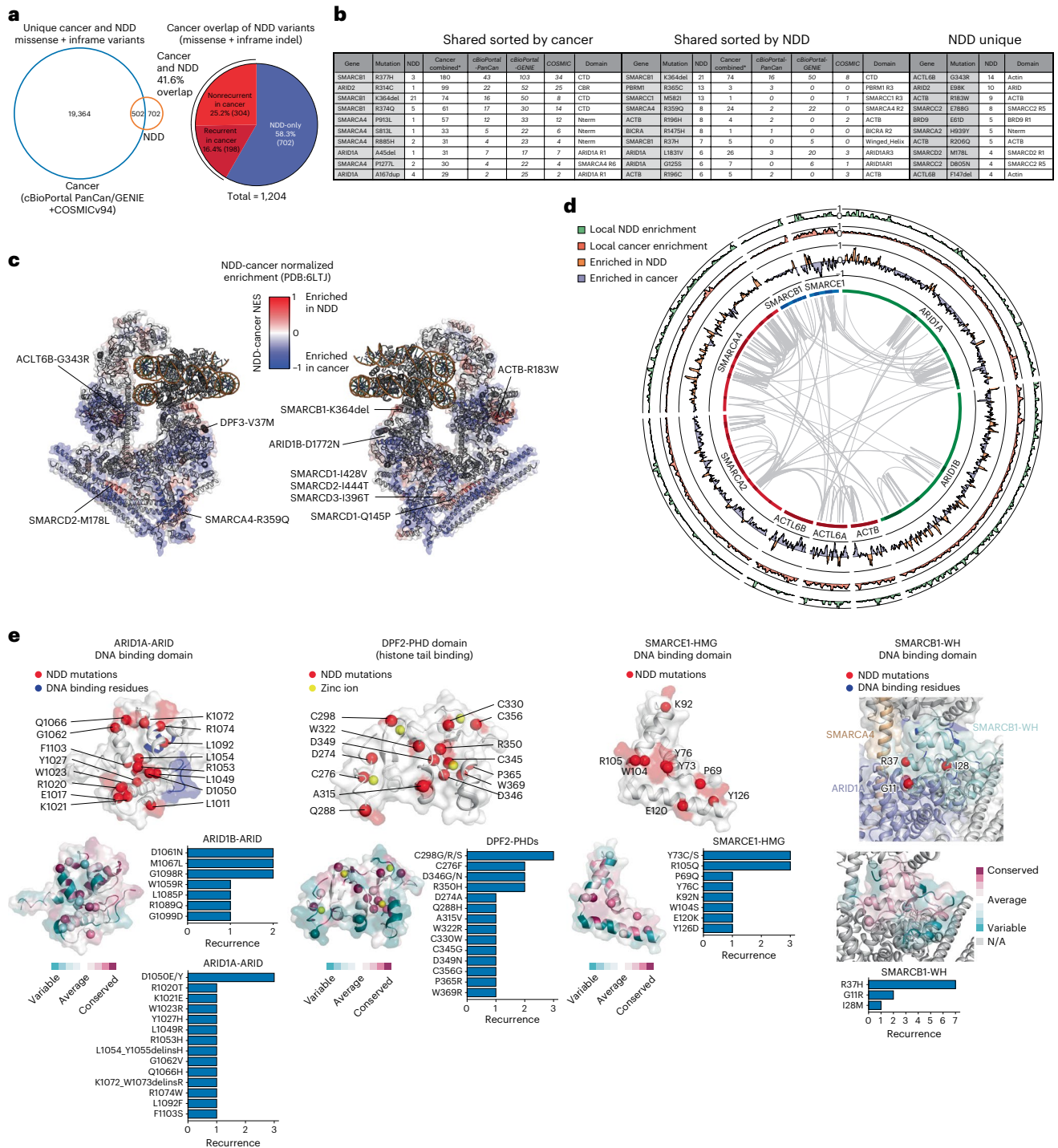
Fig. 5 | Comparison of NDD- and cancer-associated mutations in mSWI/SNF complex components. a, Venn diagram overlapping unique cancer and NDD missense and in-frame variants (left). Pie chart reflecting breakdown between NDD- and cancer-associated mSWI/SNF missense and in-frame mutations (right). The breakdown of recurrent and non-recurrent cancer variants is shown. **b**, Top ten recurrent missense and in-frame indel mutations specific to NDD and those shared between NDD and cancer, sorted by frequency in each disease type. Inter- and intradomains are indicated. **c**, Heatmap representation of mutation differences between NDD and cancer (NDD - Cancer normalized enrichment scores (Methods)) reflected on the 3D structure of the human cBAF complex (PDB:6LTJ). Red regions represent those enriched in NDD, blue represent those enriched in cancer (−1, maximally enriched in cancer; 1, maximally enriched in NDD). Labels for NDD hotspots are shown. **d**, Circos plot reflecting regions of top-mutated mSWI/SNF subunits and the local enrichment of missense and in-frame indel mutations in NDD (green), Cancer (red) or NDD-Cancer difference (represented as NDD-Cancer NES): NDD (orange) or cancer (purple); interactions between subunits, determined by cross-linking mass-spectrometry (CX-MS) performed on endogenous cBAF complexes are shown (NCP-bound endogenous cBAF, from Mashtalir et al.[93]). Scaled local recurrence, and NDD-Cancer NES were calculated similarly to panel c with one exception, where all secondary paralog mutations were preserved instead of remapping to paralogs. Enrichment scores were bounded from 0 to 1 for local recurrence and −1 to 1 for differential enrichment of mutations. Domains are represented as darker bands in the first inner ring of the Circos plot. **e**, NDD-associated mutant residues emphasized as red spheres on the structures of the ARID1A-ARID domain (PDB:1RYU), the DPF2-PHD domain (PDB:5B79), the SMARCE1-HMG DNA-binding domain (PDB:7CYU) and the SMARCB1-winged-helix DNA-binding domain (PDB:6LTJ). NDD-associated missense and inframeshift variants, ranked by frequency, are shown as bar charts. ConSurf conservation scores are mapped onto each domain structure with cyan-white-magenta color scale in increasing conservation order.
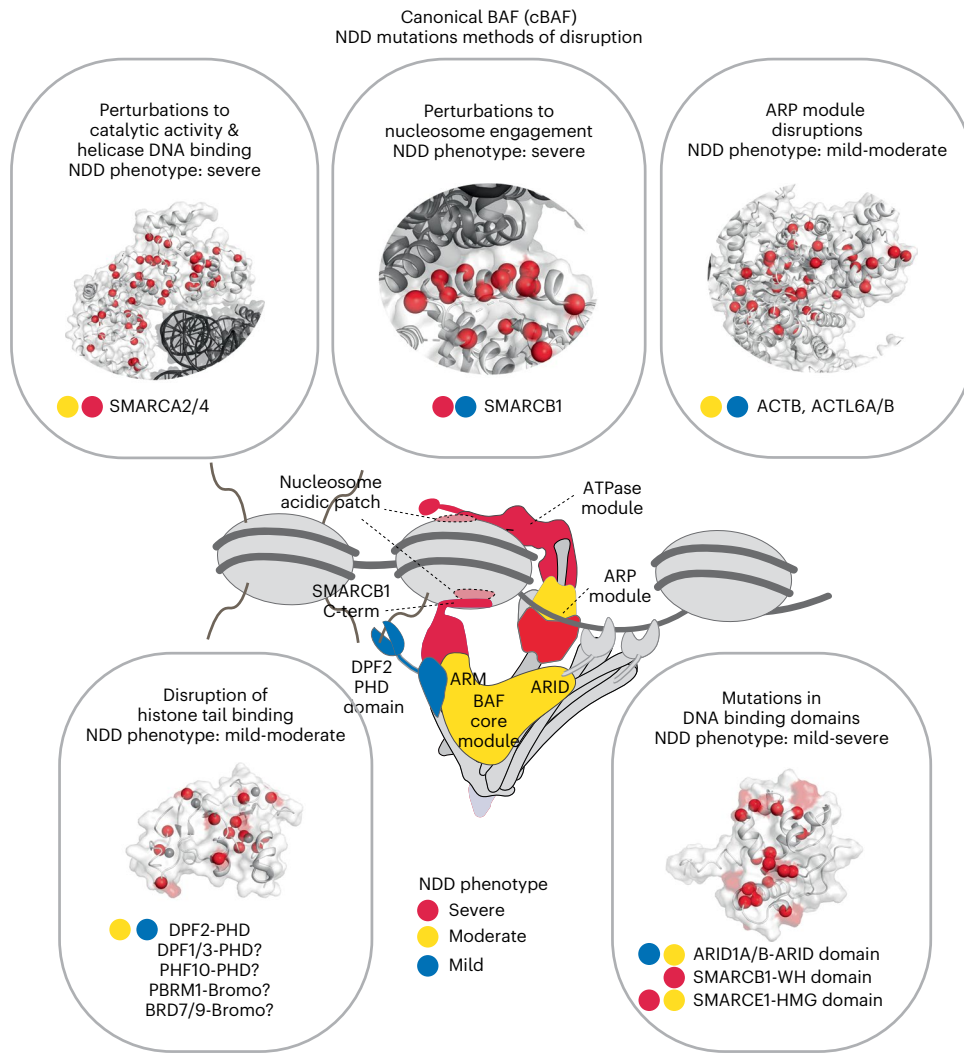
Fig. 6 | Summary of widely disrupted mSWI/SNF complex hubs in NDDs.
NDD-associated mSWI/SNF mutations occur across several subunits of the mSWI/SNF family of chromatin remodeling complexes and cluster in key structural hubs. Missense and in-frame deletions accumulate within the catalytic ATPase, nucleosome interacting, histone-binding or DNA-binding domains, as well as the ARP module, underscoring their convergence in producing neurodevelopmental aberrations. Interpretation of NDD-associated variants in the context of this framework enables mechanistic dissection of mSWI/SNF activities and provides functional links relevant to clinical phenotypes.

remodeling activity without disrupting ATPase consumption[102]. Mutation of additional nucleosome DNA-binding residues including K878, R1164 and R1142 (corresponding NDD variants: A4-K865E, A4-R1157Q/G, A2-R1105H/G/C/P/S) may have similar biochemical outcomes (Extended Data Fig. 4i). However, NDD-mutant residues in the ATP binding pocket are expected to disrupt the fundamental ATPase activity of SNF2. For example, mutation of either G797 or G795 (corresponding NDD variants A2-754A, A2-G752A and A4-G784R) residues, which provide space for ATP to bind to the ATP pocket, may reduce mSWI/SNF nucleosome remodeling activity (Extended Data Fig. 4i). Further work is required to define how mutations might impact the dynamic activity of these complexes as well as fully characterizing the structural domains not yet resolved in SMARCA2/4.

Comparing cancer and NDD mutations reveals disruption hubs
Previous studies have examined the distribution of cancer-associated single-residue mutations on the cBAF complex structure[93,94,97]. For our analysis, we examined the overlap of unique missense and inframeshift mutations identified in the context of NDD with those in human cancer (cBioPortal-PanCancer[103,104], AACR Project GENIE[105] and COSMIC[86])

(Extended Data Fig. 5a). We found that the majority (58.3%) of unique mutations found in NDD were specific to NDD (Fig. 5a, Supplementary Table 4). Further, among the 41.6% of shared cancer mutations, 16.4% were found to be recurrent among the three cancer datasets analyzed (Fig. 5a). Shared recurrent mutations in both NDD and cancer included those localized to the C-terminal domain of SMARCB1, the SMARCA4 N terminus, as well as within PBRM1 and ACTB subunits (Fig. 5b and Supplementary Table 4). By examining mutational positions rather than unique mutations, we found that over two thirds (69.3%) of NDD-mutant positions are also altered in cancer, with similar breakdown of the shared mutational recurrence (Extended Data Fig. 5b,c). Given the difficulty of de-duplicating cancer variants across the three cancer databases used in this study (cBioPortal PanCan/GENIE and COSMIC datasets), we used the cumulative recurrence across the three datasets for comparison to NDD recurrence (Fig. 5b, Extended Data Fig. 5c,d and Supplementary Table 4).

A minor positive correlation was observed between the recurrence of shared cancer (cBioPortal-PanCan) and NDD sequence variants (Extended Data Fig. 5e). Although normalization of both NDD and cancer mutational frequencies can mask regions highly mutated

in both disease settings, mutational enrichment analyses revealed several unique mutational hot spots specific to human NDD (Fig. 5c). Mutations in Arp module subunits, ACTB and ACTL6A/B, were nearly selectively enriched in NDDs, whereas mutations in the helicase domain of SMARCA4 were more enriched in cancer (Fig. 5b,c). Mutations overlapping with those in cancer localize to the SMARCA4 ATP binding pocket and nucleosomal DNA-binding residues, the SMARCB1-CTD, and the SMARCA4-BAF core module entry point (Fig. 5b and Extended Data Fig. 5f). Finally, we used cross-linking mass spectrometry (CX-MS) datasets from previous studies performed on endogenous cBAF complexes[16], which further demonstrated region-specific enrichment of NDD-versus cancer-associated mutations throughout cBAF subunits (Fig. 5d).

Mutations in structurally and functionally elusive domains
To date, 3D structural studies have resolved only ~44% of the total cBAF complex (by molecular weight), owing to the presence of low-complexity or disordered regions within many subunits (with to-date unassigned functions). Further, and given that such regions are often spaced between structured domains, several structured domains, many solved in isolation, have not been solved in the context of full 3D cBAF or PBAF complexes. We thus mapped all NDD non-truncating variants to the highly mutated ARID1A-ARID DNA-binding domain, the SMARCE1-HMG domain, the DPF2-PHD domains and the SMARCB1-WH domain to previously resolved high-resolution apo structures[106–109] (Fig. 5e and Extended Data Fig. 5g–j). Intriguingly, the majority of ARID1A-ARID domain and SMARCB1 WH domain non-truncating variants do not overlap with the DNA-binding residues, and we therefore predict that they disrupt intradomain structural integrity (Fig. 5e and Extended Data Fig. 5g–h)[108]. As has been demonstrated previously, mutations in the DPF2-PHD domains disrupt zinc-binding residues which are important for PHD domain structural formation, resulting in decreased affinity to modified histone substrates (Fig. 5e and Extended Data Fig. 5i)[109]. NDD-associated mutations in the SMARCE1-HMG domain accumulate on the DNA-binding interface of the structure (Fig. 5e and Extended Data Fig. 5j)[107] and hence are predicted to inhibit DNA binding.

Discussion
Here, we demonstrate that mSWI/SNF complex genes are the most frequently disrupted chromatin regulatory entity in NDD, with perturbation of several key structural 'hubs' within this multicomponent complex displaying a phenotypic convergence that yields NDD features associated in the literature with the greatest level of NDD severity (Fig. 1d and Fig. 6). Our study serves as a powerful foundation upon which to pursue integrated efforts between the chromatin biology and neurobiology communities to functionally characterize and prioritize these frequent disruptions.

It should be noted that because the products of mSWI/SNF complex genes are assembled into a highly heterogeneous group of complexes, the total extent of mutational burden of this complex reported here may not be completely recognized, even with genes such as *ARID1B* ranking among the most highly mutated in NDDs (Extended Data Fig. 1n)[110–112]. Disruption of both structured and unstructured domains presented here may impart altered mSWI/SNF complex localization and activity on the genome via a range of mechanisms requiring extensive further investigation. Additionally, further examination of zygosity and how missense variants within the same protein differentially impact protein activity may reveal distinct functions. For example, both dominant and recessive single amino acid variants affected *ACTL6B* have been identified[113]. Although the ACTL6B G393R recessive variant has been shown to reduce ACTL6B protein expression, behaving as a loss-of-function mutation[114], the dominant G343R variant is predicted to impart dominant-negative effects that disrupting mSWI/SNF activity[42].

In this study, we curated a list of chromatin regulatory genes in combination with the EpiFactor database to investigate the prevalence of chromatin-related process disruptions in NDD. However, additional work is needed to define a maximally complete set of chromatin regulators, regulatory complexes and their subunit membership. Further, functional studies must be performed to define mechanisms by which variants alter activity or other functions, especially given that 3D structures are based on a range of complex states and conformations, which may vary in biologic relevance. Importantly, although we have obtained information on recurrence of sequence variants for which distinct cases were clear, potential duplicates were omitted in processing in cases for which we could not verify distinct cases between literature and databases used, meaning that recurrence of some variants may be artificially reduced. Further, cross-referencing of additional private databases such as FoundationCORE may be useful in follow-up analyses[115]. To prevent inclusion of false positives, we omitted NDD-associated mSWI/SNF sequence variants which are also present in gnomAD with a minor allele frequency of >0.5%, predicted to be benign. Although the overwhelming majority (96%) of DECIPHER variants reported to date are heterozygous (Extended Data Fig. 1l), zygosity data were not included in this study, and this remains a limitation. By centering the majority of our analysis on de novo variants, we expect these to be pathogenic; however, future studies must be performed to assess the full scope of the molecular and pathophysiological consequences of these mutations.

Online content
Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-023-01451-6.

References

1. Bailey, M. H. et al. Comprehensive characterization of cancer driver genes and mutations. *Cell* **173**, 371–385 (2018).
2. Gabriele, M., Tobon, A. L., D'Agostino, G. & Testa, G. The chromatin basis of neurodevelopmental disorders: Rethinking dysfunction along the molecular and temporal axes. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **84**, 306–327 (2018).
3. Rubeis, S. D. et al. Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* **515**, 209–215 (2014).
4. Satterstrom, F. K. et al. Large-scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism. *Cell* **180**, 568–584 (2020).
5. Clapier, C. R., Iwasa, J., Cairns, B. R. & Peterson, C. L. Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nat. Rev. Mol. Cell Biol.* **18**, 407–422 (2017).
6. Sundaramoorthy, R. & Owen-Hughes, T. Chromatin remodelling comes into focus. *F1000Res.* **9**, Faculty Rev-1011 (2020).
7. Valencia, A. M. & Kadoch, C. Chromatin regulatory mechanisms and therapeutic opportunities in cancer. *Nat. Cell Biol.* **21**, 152–161 (2019).
8. Kadoch, C. & Crabtree, G. R. Mammalian SWI/SNF chromatin remodeling complexes and cancer: Mechanistic insights gained from human genomics. *Sci. Adv.* **1**, e1500447 (2015).
9. Kadoch, C. et al. Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. *Nat. Genet.* **45**, 592–601 (2013).
10. Maulik, P. K., Mascarenhas, M. N., Mathers, C. D., Dua, T. & Saxena, S. Prevalence of intellectual disability: A meta-analysis of population-based studies. *Res. Dev. Disabil.* **32**, 419–436 (2011).

11. Robertson, J., Hatton, C., Emerson, E. & Baines, S. Prevalence of epilepsy among people with intellectual disabilities: A systematic review. *Seizure* **29**, 46–62 (2015).

12. Kleefstra, T., Schenck, A., Kramer, J. M. & van Bokhoven, H. The genetics of cognitive epigenetics. *Neuropharmacology* **80**, 83–94 (2014).

13. Ronan, J. L., Wu, W. & Crabtree, G. R. From neural development to cognition: Unexpected roles for chromatin. *Nat. Rev. Genet.* **14**, 347–359 (2013).

14. Valencia, A. M. & Paşca, S. P. Chromatin dynamics in human brain development and disease. *Trends Cell Biol.* **32**, 98–101 (2021).

15. Kelso, T. W. R. et al. Chromatin accessibility underlies synthetic lethality of SWI/SNF subunits in ARID1A-mutant cancers. *Elife* **6**, e30506 (2017).

16. Mashtalir, N. et al. Modular organization and assembly of SWI/SNF family chromatin remodeling complexes. *Cell* **175**, 1272–1288.e20 (2018).

17. Nakayama, R. T. et al. SMARCB1 is required for widespread BAF complex-mediated activation of enhancers and bivalent promoters. *Nat. Genet.* **49**, 1613–1623 (2017).

18. Pan, J. et al. The ATPase module of mammalian SWI/SNF family complexes mediates subcomplex identity and catalytic activity-independent genomic targeting. *Nat. Genet.* **51**, 618–626 (2019).

19. Santen, G. W. E. et al. Mutations in SWI/SNF chromatin remodeling complex gene ARID1B cause Coffin-Siris syndrome. *Nat. Genet.* **44**, 379–380 (2012).

20. Tsurusaki, Y. et al. Mutations affecting components of the SWI/SNF complex cause Coffin-Siris syndrome. *Nat. Genet.* **44**, 376–378 (2012).

21. Wieczorek, D. et al. A comprehensive molecular study on Coffin–Siris and Nicolaides–Baraitser syndromes identifies a broad molecular and clinical spectrum converging on altered chromatin remodeling. *Hum. Mol. Genet.* **22**, 5121–5135 (2013).

22. Firth, H. V. et al. DECIPHER: Database of chromosomal imbalance and phenotype in humans using Ensembl resources. *Am. J. Hum. Genet.* **84**, 524–533 (2009).

23. Landrum, M. J. et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.* **46**, D1062–D1067 (2018).

24. Fokkema, I. F. A. C. et al. LOVD v.2.0: the next generation in gene variant databases. *Hum. Mutat.* **32**, 557–563 (2011).

25. An, J.-Y. et al. Genome-wide de novo risk score implicates promoter variation in autism spectrum disorder. *Science* **362**, eaat6576 (2018).

26. Appenzeller, S. et al. De novo mutations in synaptic transmission genes including DNM1 cause epileptic encephalopathies. *Am. J. Hum. Genet.* **95**, 360–370 (2014).

27. Gilissen, C. et al. Genome sequencing identifies major causes of severe intellectual disability. *Nature* **511**, 344–347 (2014).

28. Halldorsson, B. V. et al. Characterizing mutagenic effects of recombination through a sequence-level genetic map. *Science* **363**, eaau1043 (2019).

29. Homsy, J. et al. De novo mutations in congenital heart disease with neurodevelopmental and other congenital anomalies. *Science* **350**, 1262–1266 (2015).

30. Iossifov, I. et al. De novo gene disruptions in children on the autistic spectrum. *Neuron* **74**, 285–299 (2012).

31. Iossifov, I. et al. The contribution of de novo coding mutations to autism spectrum disorder. *Nature* **515**, 216–221 (2014).

32. Jin, S. C. et al. Contribution of rare inherited and de novo variants in 2,871 congenital heart disease probands. *Nat. Genet* **49**, 1593–1601 (2017).

33. Kaplanis, J. et al. Evidence for 28 genetic disorders discovered by combining healthcare and research data. *Nature* **586**, 757–762 (2020).

34. McRae, J. F. et al. Prevalence and architecture of de novo mutations in developmental disorders. *Nature* **542**, 433–438 (2017).

35. Lelieveld, S. H. et al. Meta-analysis of 2,104 trios provides support for 10 new genes for intellectual disability. *Nat. Neurosci.* **19**, 1194–1196 (2016).

36. Ligt, Jde et al. Diagnostic exome sequencing in persons with severe intellectual disability. *N. Engl. J. Med.* **367**, 1921–1929 (2012).

37. Sanders, S. J. et al. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* **485**, 237–241 (2012).

38. Rauch, A. et al. Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. *Lancet* **380**, 1674–1682 (2012).

39. Sanders, S. J. et al. Insights into autism spectrum disorder genomic architecture and biology from 71 risk loci. *Neuron* **87**, 1215–1233 (2015).

40. Alvarez-Mora, M. I. et al. Comprehensive molecular testing in patients with high functioning autism spectrum disorder. *Mutat. Res Fundam. Mol. Mech. Mutagen* **784**, 46–52 (2016).

41. Aref-Eshghi, E. et al. BAFopathies' DNA methylation epi-signatures demonstrate diagnostic utility and functional continuum of Coffin–Siris and Nicolaides–Baraitser syndromes. *Nat. Commun.* **9**, 4885 (2018).

42. Bell, S. et al. Mutations in ACTL6B Cause Neurodevelopmental Deficits and Epilepsy and Lead to Loss of Dendrites in Human Neurons. *Am. J. Hum. Genet.* **104**, 815–834 (2019).

43. Bowling, K. M. et al. Genomic diagnosis for children with intellectual disability and/or developmental delay. *Genome Med.* **9**, 43 (2017).

44. Yuen, R. K. C. et al. Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. *Nat. Neurosci.* **20**, 602–611 (2017).

45. Campeau, P. M. & Hennekam, R. C., group, T. D. syndrome collaborative. DOORS syndrome: Phenotype, genotype and comparison with Coffin-Siris syndrome. *Am. J. Med. Genet. Part C. Semin. Med. Genet.* **166**, 327–332 (2014).

46. Chérot, E. et al. Using medical exome sequencing to identify the causes of neurodevelopmental disorders: Experience of 2 clinical units and 216 patients. *Clin. Genet.* **93**, 567–576 (2018).

47. D'Gama, A. M. et al. Targeted DNA Sequencing from Autism Spectrum Disorder Brains Implicates Multiple Genetic Mechanisms. *Neuron* **88**, 910–917 (2015).

48. Diets, I. J. et al. A recurrent de novo missense pathogenic variant in SMARCB1 causes severe intellectual disability and choroid plexus hyperplasia with resultant hydrocephalus. *Genet. Med.* **21**, 572–579 (2019).

49. Doan, R. N. et al. Mutations in Human Accelerated Regions Disrupt Cognition and Social Behavior. *Cell* **167**, 341–354.e12 (2016).

50. Farwell, K. D. et al. Enhanced utility of family-centered diagnostic exome sequencing with inheritance model-based analysis: results from 500 unselected families with undiagnosed genetic conditions. *Genet. Med.* **17**, 578–586 (2015).

51. Fichera, M. et al. Mutations in ACTL6B, coding for a subunit of the neuron-specific chromatin remodeling complex nBAF, cause early onset severe developmental and epileptic encephalopathy with brain hypomyelination and cerebellar atrophy. *Hum. Genet.* **138**, 187–198 (2019).

52. Geisheker, M. R. et al. Hotspots of missense mutation identify novel neurodevelopmental disorder genes and functional domains. *Nat. Neurosci.* **20**, 1043–1051 (2017).

53. Guo, H. et al. Genome sequencing identifies multiple deleterious variants in autism patients with more severe phenotypes. *Genet. Med.* **21**, 1611–1620 (2019).

54. Karaca, E. et al. Genes that affect brain structure and function identified by rare variant analyses of Mendelian neurologic disease. *Neuron* **88**, 499–513 (2015).

55. Kleefstra, T. et al. Disruption of an EHMT1-associated chromatin-modification module causes intellectual disability. *Am. J. Hum. Genet.* **91**, 73–82 (2012).

56. Koga, M. et al. Involvement of SMARCA2/BRM in the SWI/SNF chromatin-remodeling complex in schizophrenia. *Hum. Mol. Genet* **18**, 2483–2494 (2009).

57. Krumm, N. et al. Excess of rare, inherited truncating mutations in autism. *Nat. Genet.* **47**, 582–588 (2015).

58. Lecoquierre, F. et al. Variant recurrence in neurodevelopmental disorders: the use of publicly available genomic data identifies clinically relevant pathogenic missense variants. *Genet. Med.* **21**, 2504–2511 (2019).

59. Li, J. et al. Targeted sequencing and functional analysis reveal brain-size-related genes and their networks in autism spectrum disorders. *Mol. Psychiatr.* **22**, 1282–1290 (2017).

60. Consortium, A. S. et al. Rates, distribution and implications of postzygotic mosaic mutations in autism spectrum disorder. *Nat. Neurosci.* **20**, 1217–1224 (2017).

61. Machol, K. et al. Expanding the spectrum of BAF-related disorders: De novo variants in SMARCC2 cause a syndrome with intellectual disability and developmental delay. *Am. J. Hum. Genet.* **104**, 164–178 (2019).

62. Mannino, E. A., Miyawaki, H., Santen, G. & Vergano, S. A. S. First data from a parent-reported registry of 81 individuals with Coffin–Siris syndrome: Natural history and management recommendations. *Am. J. Med. Genet. A* **176**, 2250–2258 (2018).

63. Marom, R. et al. Heterozygous variants in ACTL6A, encoding a component of the BAF complex, are associated with intellectual disability. *Hum. Mutat.* **38**, 1365–1371 (2017).

64. Mignot, C. et al. ARID1B mutations are the major genetic cause of corpus callosum anomalies in patients with intellectual disability. *Brain* **139**, e64 (2016).

65. Monies, D. et al. Lessons learned from large-scale, first-tier clinical exome sequencing in a highly consanguineous population. *Am. J. Hum. Genet.* **104**, 1182–1201 (2019).

66. Nixon, K. C. J. et al. A syndromic neurodevelopmental disorder caused by mutations in SMARCD1, a core SWI/SNF subunit needed for context-dependent neuronal gene regulation in flies. *Am. J. Hum. Genet.* **104**, 596–610 (2019).

67. Pascolini, G., Agolini, E., Novelli, A., Majore, S. & Grammatico, P. The p.Arg377Trp variant in ACTL6A underlines a recognizable BAF-opathy phenotype. *Clin. Genet* **97**, 672–674 (2020).

68. Sandestig, A. et al. Could dissimilar phenotypic effects of ACTB missense mutations reflect the actin conformational change two novel mutations and literature review. *Mol. Syndromol.* **9**, 259–265 (2019).

69. Santen, G. W. E. et al. Coffin–Siris syndrome and the BAF complex: Genotype–phenotype study in 63 patients. *Hum. Mutat.* **34**, 1519–1528 (2013).

70. Sekiguchi, F. et al. Genetic abnormalities in a large cohort of Coffin–Siris syndrome patients. *J. Hum. Genet.* **64**, 1173–1186 (2019).

71. Tsurusaki, Y. et al. Coffin–Siris syndrome is a SWI/SNF complex disorder. *Clin. Genet.* **85**, 548–554 (2014).

72. Zarate, Y. A. et al. SMARCE1, a rare cause of Coffin–Siris syndrome: Clinical description of three additional cases. *Am. J. Med Genet. A* **170**, 1967–1973 (2016).

73. Houdt, J. K. J. V. et al. Heterozygous missense mutations in SMARCA2 cause Nicolaides-Baraitser syndrome. *Nat. Genet.* **44**, 445–449 (2012).

74. Vasileiou, G. et al. Mutations in the BAF-complex subunit DPF2 are associated with Coffin-Siris syndrome. *Am. J. Hum. Genet.* **102**, 468–479 (2018).

75. Verloes, A. et al. Baraitser–Winter cerebrofrontofacial syndrome: Delineation of the spectrum in 42 cases. *Eur. J. Hum. Genet.* **23**, 292–301 (2015).

76. Vissers, L. E. L. M. et al. A clinical utility study of exome sequencing versus conventional genetic testing in pediatric neurology. *Genet. Med.* **19**, 1055–1063 (2017).

77. Wolff, D. et al. In-frame deletion and missense mutations of the C-terminal helicase domain of SMARCA2 in three patients with Nicolaides-Baraitser syndrome. *Mol. Syndromol.* **2**, 237–244 (2012).

78. Wu, H. et al. Phenotype-to-genotype approach reveals head-circumference-associated genes in an autism spectrum disorder cohort. *Clin. Genet* **97**, 338–346 (2020).

79. Xiong, J. et al. Neurological diseases with autism spectrum disorder: Role of ASD risk genes. *Front. Neurosci.* **13**, 349 (2019).

80. Yu, Y. et al. De novo mutations in ARID1B associated with both syndromic and non-syndromic short stature. *Bmc Genomics* **16**, 701 (2015).

81. Zhao, J. J. et al. Exome sequencing reveals NAA15 and PUF60 as candidate genes associated with intellectual disability. *Am. J. Med. Genet.* **177**, 10–20 (2018).

82. Medvedeva, Y. A. et al. EpiFactors: A comprehensive database of human epigenetic factors and complexes. *Database (Oxford)* **2015**, bav067 (2015).

83. Koopmans, F. et al. SynGO: An evidence-based, expert-curated knowledge base for the synapse. *Neuron* **103**, 217–234 (2019).

84. Abrahams, B. S. et al. SFARI Gene 2.0: A community-driven knowledgebase for the autism spectrum disorders (ASDs). *Mol. Autism* **4**, 36 (2013).

85. Bögershausen, N. & Wollnik, B. Mutational landscapes and phenotypic spectrum of SWI/SNF-related intellectual disability disorders. *Front Mol. Neurosci.* **11**, 252 (2018).

86. Tate, J. G. et al. COSMIC: The catalogue of somatic mutations In cancer. *Nucleic Acids Res.* **47**, D941–D947 (2019).

87. Karczewski, K. J. et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).

88. Cappuccio, G. et al. De novo SMARCA2 variants clustered outside the helicase domain cause a new recognizable syndrome with intellectual disability and blepharophimosis distinct from Nicolaides–Baraitser syndrome. *Genet. Med.* **22**, 1838–1850 (2020).

89. Sluijs, P. Jvander et al. Discovering a new part of the phenotypic spectrum of Coffin-Siris syndrome in a fetal cohort. *Genet. Med.* **24**, 1753–1760 (2022).

90. van der Sluijs, P. J. et al. A case series of familial ARID1B variants illustrating variable expression and suggestions to update the ACMG criteria. *Genes (Basel)* **12**, 1275 (2021).

91. Milone, R., Gnazzo, M., Stefanutti, E., Serafin, D. & Novelli, A. A new missense mutation in DPF2 gene related to Coffin Siris syndrome 7: Description of a mild phenotype expanding DPF2-related clinical spectrum and differential diagnosis among similar syndromes epigenetically determined. *Brain Dev.* **42**, 192–198 (2019).

92. Adzhubei, I. A. et al. A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).

93. Mashtalir, N. et al. A structural model of the endogenous human BAF complex informs disease mechanisms. *Cell* **183**, 802–817.e24 (2020).

94. He, S. et al. Structure of nucleosome-bound human BAF complex. *Science* **367**, 875–881 (2020).

95. Donato, N. D. et al. Severe forms of Baraitser–Winter syndrome are caused by ACTB mutations rather than ACTG1 mutations. *Eur. J. Hum. Genet.* **22**, 179–183 (2014).

96. Ittisoponpisan, S. et al. Can predicted protein 3D structures provide reliable insights into whether missense variants are disease associated? *J. Mol. Biol.* **431**, 2197–2212 (2019).

97. Yuan, J., Chen, K., Zhang, W. & Chen, Z. Structure of human chromatin-remodelling PBAF complex bound to a nucleosome. *Nature* **605**, 166–171 (2022).

98. Valencia, A. M. et al. Recurrent SMARCB1 mutations reveal a nucleosome acidic patch interaction site that potentiates mSWI/SNF complex chromatin remodeling. *Cell* **179**, 1342–1356. e23 (2019).

99. Khanna, T., Hanna, G., Sternberg, M. J. E. & David, A. Missense3D-DB web catalogue: An atom-based analysis and repository of 4M human protein-coding genetic variants. *Hum. Genet.* **140**, 805–812 (2021).

100. Han, Y., Reyes, A. A., Malik, S. & He, Y. Cryo-EM structure of SWI/SNF chromatin remodeling complex with nucleosome. *Nature* **579**, 452–455 (2020).

101. Liu, X., Li, M., Xia, X., Li, X. & Chen, Z. Mechanism of chromatin remodelling revealed by the Snf2-nucleosome structure. *Nature* **544**, 440–445 (2017).

102. Li, M. et al. Mechanism of DNA translocation underlying chromatin remodelling by Snf2. *Nature* **567**, 409–413 (2019).

103. Gao, J. et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal* **6**, pl1 (2013).

104. Cerami, E. et al. The cBio Cancer Genomics Portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2**, 401–404 (2012).

105. The AACR Project GENIE Consortium. AACR Project GENIE: Powering precision medicine through an international consortium. *Cancer Discov.* **7**, 818–831 (2017).

106. Allen, M. D., Freund, S. M. V., Zinzalla, G. & Bycroft, M. The SWI/SNF subunit INI1 contains an N-terminal winged helix DNA binding domain thatis a target for mutations in schwannomatosis. *Struct. Lond. Engl.* **1993**, 1344–1349 (2015).

107. Heo, Y. et al. Crystal structure of the HMG domain of human BAF57 and its interaction with four-way junction DNA. *Biochem. Biophys. Res. Commun.* **533**, 919–924 (2020).

108. Kim, S., Zhang, Z., Upchurch, S., Isern, N. & Chen, Y. Structure and DNA-binding sites of the SWI1 AT-rich interaction domain (ARID) suggest determinants for sequence-specific DNA recognition. *J. Biol. Chem.* **279**, 16670–16676 (2004).

109. Xiong, X. et al. Selective recognition of histone crotonylation by double PHD fingers of MOZ and DPF2. *Nat. Chem. Biol.* **12**, 1111–1118 (2016).

110. Hoyer, J. et al. Haploinsufficiency of ARID1B, a member of the SWI/SNF-A chromatin-remodeling complex, is a frequent cause of intellectual disability. *Am. J. Hum. Genet.* **90**, 565–572 (2012).

111. Sluijs, P. Jvander et al. The ARID1B spectrum in 143 patients: From nonsyndromic intellectual disability to Coffin–Siris syndrome. *Genet. Med.* **21**, 1295–1307 (2019).

112. Wright, C. F. et al. Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data. *Lancet Lond. Engl.* **385**, 1305–1314 (2015).

113. Rowland, M. E., Jajarmi, J. M., Osborne, T. S. M. & Ciernia, A. V. Insights Into the emerging role of Baf53b in autism spectrum disorder. *Front. Mol. Neurosci.* **15**, 805158 (2022).

114. Wenderski, W. et al. Loss of the neural-specific BAF subunit ACTL6B relieves repression of early response genes and causes recessive autism. *Proc. Natl Acad. Sci. USA* **117**, 10055–10066 (2020).

115. Hartmaier, R. J. et al. High-throughput genomic profiling of adult solid tumors reveals novel insights into cancer pathogenesis. *Cancer Res.* **77**, 2464–2475 (2017).

Methods

Novel variant collection

Novel NDD-related mSWI/SNF gene variants reported in this study were identified through physician referrals and the Coffin-Siris syndrome registry. Variants from Leiden University Medical Center were identified in a diagnostic setting, and genetic data were retrieved from the generated reports or shared with us by the treating physician with consent from the patient or parents. The institutional review board of Leiden University Medical Center provided approval waivers for using de-identified data and publishing aggregated data (G18.098 and G21.129) without obtaining specific informed consent. Individuals identified through Eastern Virginia Medical School were recruited to the Coffin-Siris syndrome registry through clinicians, social media and patient foundations. Individuals completed an online consent form followed by a registry survey with phenotypic questions. The Coffin-Siris Syndrome Registry has been approved by the Eastern Virginia Medical School institutional review board (15-03-EX-0058). Novel variants reported in this study have been deposited in LOVD (https://www.lovd.nl/)[24]. Variants identified through this method that were present in previously published literature or deposited in an online repository were excluded for analysis in this study to prevent reporting potential duplicates (Curating mSWI/SNF gene NDD-associated variants section). Given that our paper centers on the mutational rather than phenotypic outcomes of NDD-related mSWI/SNF variants, future clinical papers will further explore the phenotypes associated with novel variants published in this manuscript. During the review process, some novel variants included in this study were published with detailed clinical information[89].

Mutational datasets

Open-access mutations publicly available on the DECIPHER database (https://www.deciphergenomics.org/; accessed June 22, 2022) (ref. 22) were used for broader chromatin gene analysis (Fig. 1c,d and Extended Data Fig. 1k,l). The queried chromatin remodeling complex gene list (SWI/SNF, CHD, INO80 and ISWI) was manually curated from a literature review detailed below (Supplementary Table 2).

Chromatin regulatory gene sets (Supplementary Table 2)

Chromatin remodeling complex gene lists were curated from a variety of sources, including HGNC gene groups SWI/SNF and INO80 (https://www.genenames.org/data/genegroup/#!/), as well as a literature review of all chromatin remodeling complexes[116,117], mSWI/SNF[16], ISWI[118], CHD[119] and INO80 (refs. 120–124). The histone modifier gene list was gathered from *HISTome2* (refs. 125,126) (http://www.actrec.gov.in/histome2/). Polycomb repressive complex genes and DNA methylation regulatory genes were informed by the literature[127,128]. Additional chromatin regulatory complexes were obtained from EpiFactor[82] (https://epifactors.autosome.org/protein_complexes). The full set of cBAF, PBAF and ncBAF genes were included in the EpiFactor complexes if absent.

Curating mSWI/SNF gene NDD-associated variants

The set of rare inherited and de novo variants included data from three cohorts of individuals with autism spectrum disorders or other developmental disorders: the Simons SSC/ASC, SPARK and DDD cohorts. Details about merging and de-duplicating the data are described in Fu et al.[129] Briefly, duplicated samples were identified and excluded by IBD and other metadata, and the filtered samples were merged to provide a single unified set of de-duplicated de novo variants in autism spectrum disorders and other developmental disorders. The recurrence of NDD de novo variants across BAF genes and several gene sets of interest, including a curated set of chromatin remodelers, epigenetic modifiers and synaptic genes were visualized with scatter plots and bar charts using matplotlib[130]. The set of de novo variants and non-benign SNVs in DECIPHER were used for all summary calculations in Fig. 1 and Extended Data Fig. 1 and for comparisons between the BAF genes,

chromatin regulatory genes, epigenetic modifier genes and synaptic genes. The queried chromatin regulatory gene list was based on EpiFactor (https://epifactors.autosome.ru/genes; accessed 2 September 2021) (ref. 82 updated to include all mSWI/SNF genes (Supplementary Table 2). The queried synaptic gene list was based on the SynGO gene list (https://www.syngoportal.org/; accessed 2 September 2021) (ref. 83). The development disorder DECIPHER gene list was based on DDG2P genes in DECIPHER (accessed 13 June 2022).

A comprehensive list of SNV and short in-frame indels (inframeshift variants) was compiled from an extensive literature review, the combined set of rare inherited and de novo variants from the Simons SSC/ASC, SPARK, and DDD cohorts (the 'combined cohort study'), the DECIPHER database of SNVs (https://www.deciphergenomics.org/), the merged set of de novo mutations from the DNM effort by McRae et al.[34] NDD-associated ClinVar mutations (accessed 5/15/2021), NDD-associated variants from LOVD (LOVD v3.0 accessed June 2022) and 85 previously unreported cases published in this study collected through the laboratories of S.A.S.V. (Eastern Virginia Medical School) and G.W.E.S. (Leiden University Medical Center).

First, the combined set of rare inherited and de novo variants was split into a set of rare inherited variants and a set of de novo variants. All rare inherited PTVs, in-frame indel variants and de novo variants were included in the integrated dataset. Guided by the analysis in Fu et al.[129], where missense variants with MPC scores (missense badness, PolyPhen-2 and constraint) of 1 or more were observed to confer moderate to strong levels of risk in developing autism and missense rare inherited variants with MPC scores ≥1 were included in the integrated dataset. All other rare inherited variants from the combined cohort study were excluded. Then, samples were cross-referenced between the combined cohort study, DECIPHER database, and the DNM cohort of de novo mutations and identical variants from the same samples (using available sample IDs or aliases) were removed to de-deduplicate the data between these three cohorts/databases. Separately, a list of de novo variants in BAF genes across several other studies in the literature not covered previously by the cohorts used in DECIPHER and the combined cohort study (SSC/ASC, SPARK and DDD) were manually curated and de-duplicated to form the compiled set of mutations from the literature. Additionally, NDD-associated mutations from the LOVD database were compiled and filtered to include all PTV and in-frame indels and de novo/likely de novo missense variants. All benign/likely benign variants were excluded. The filtered set of LOVD variants and the manually curated variants from the literature were merged and de-duplicated based on sample IDs or aliases (if available) and study ID / reference (if sample IDs were not available). For shared variants between LOVD and the literature, where it was not clear whether these variants were duplicates, only shared variants from the manually curated literature dataset were kept, effectively de-duplicating the data. Minimal overlap was assumed between the de-duplicated set of LOVD/literature variants and the de-duplicated set of SSC + ASC/SPARK/DDD/DECIPHER/DNM variants. These two sets were merged, followed by a round of manual curation to double check that as many duplicates or potential duplicates were removed during dataset integration. The set of 85 novel cases identified by S.A.S.V. and G.W.E.S. were added to this merged dataset. In parallel, a curated set of ClinVar variants from samples with NDD-associated clinical features and unknown/likely pathogenic/pathogenic clinical significance was generated. Benign and likely benign ClinVar variants were excluded. Additionally, ClinVar variants submitted by GeneDx were excluded due to substantial overlap with the comprehensive analysis of de novo mutations in NDD by Kaplanis et al. included in the DNM database of de novo mutations. Samples were de-duplicated between ClinVar and the LOVD/literature dataset using SCV codes wherever available. Finally, this de-duplicated ClinVar dataset was used to adjust the counts of the previously merged dataset of NDD-associated BAF mutations from the combined cohort study (SSC/ASC, SPARK

and DDD), DECIPHER SNVs, DNM, LOVD and the literature. It was difficult (and sometimes impossible) to track, match and assign each filtered NDD-associated ClinVar SCV (submitted record for each variant) with the list of available sample IDs or aliases in the previously merged dataset. Thus, the total counts for each variant were adjusted to the total counts found in ClinVar (based on the number of submissions for each variant using SCV IDs) to eliminate the possibility of double counting if the ClinVar total count for a variant was more than the total count from the previously merged dataset. This procedure assumes submissions to ClinVar overlap entirely with the previously merged dataset, so it is possible the new merged dataset containing ClinVar variants might undercount some NDD-associated BAF variants. This integrated dataset was compared to gnomAD v3.1.2 to remove potential SNPs and other variants that occur frequently in a collection of healthy individuals. A more stringent MAF threshold of ≥0.5% MAF was used to exclude potentially common variants in the integrated dataset. This final integrated dataset was manually checked once more to exclude potential duplicates and likely benign variants before freezing for all downstream analyses. A total of 2,539 NDD-associated BAF variants are included in this dataset, including 85 novel cases and 72 previously unreported variants.

To standardize the data, all variants were remapped to the UniProt canonical BAF protein isoforms (see Supplementary Table 3), and duplicates that could not be confirmed unique cases were removed. Unless otherwise noted, remapping of all variants (both NDD variants and cancer variants) to different isoforms was performed using the Ensembl Variant Effect Predictor (VEP) online web server[131].

gnomAD variants of the general population were derived from the gnomAD v3 dataset (accessed 11 January 2021).

Cancer dataset cleaning and compilation

PanCancer datasets from TCGA and cBioPortal[103,104] were cleaned and compiled for all downstream analyses related to NDD versus cancer comparisons.

The TCGA MC3 PanCancer dataset was used for NDD versus cancer comparisons in Extended Data Fig. 1. Briefly, known SNPs were removed and BAF gene mutations were remapped to the canonical UniProt transcripts (Supplementary Table 3). Missense, nonsense and frameshift mutations were included, and all other mutations were excluded. This filtered set of mutations merged with the combined cohort study of NDD-associated mutations from the combined SSC/ASC, SPARK and DDD cohorts. Total cancer missense, frameshift and nonsense mutational recurrence was log normalized, compared to total de novo NDD-associated missense and PTV mutational recurrence for each gene, and visualized as a scatterplot using matplotlib[130], with BAF genes indicated in red. The total proportion of NDD and Cancer missense and PTV mutations across the BAF genes were visualized as a grouped bar chart using matplotlib[130].

Mutations across BAF genes from the curated set of nonredundant studies in cBioPortal, the AACR Project GENIE (accessed through cBioPortal) and COSMIC were compiled and filtered for NDD versus cancer comparative analyses across the BAF genes. Briefly, the BAF mutations were remapped to the UniProt canonical BAF protein isoforms (Supplementary Table 3) using the Ensembl VEP online web server[131]. Missense, frameshift, nonsense and in-frame indels were included, and all other mutations were excluded. Additionally, duplicate mutations in patients with multiple samples were excluded. This filtered set of mutations from cBioPortal[103,104] was used for downstream BAF cancer versus NDD comparative analyses.

NDD gene set enrichment analysis

A custom Perl[132] script was used to determine the enrichment of GOMF gene sets enriched in DDG2P genes, a list of genes known to be associated with developmental disorders. All BAF genes were added back to DDG2P gene list if absent. Specifically, GOMF gene sets were overlapped with DDG2P using gene symbols and a hypergeometric distribution test (for example, statistical overrepresentation test) was used to evaluate the significance (P value) of enrichment of each GOMF. Additionally, the total and mean number of de novo missense and PTVs in ASD + DD using the combined cohort study was calculated for the overlapping genes (using gene symbols) between each GOMF gene set and DDG2P genes. The enrichment of GOMFs in DDG2P genes were visualized as scatterplots and ranked by significance (P value) and total de novo missense and PTV mutational recurrence for the overlapping genes (using gene symbols) with the top 10 GOMFs labeled. Additionally, the top 50 most enriched GOMFs by statistical significance (P value) were ranked by the mean number of de novo missense and PTVs in the overlapping genes (using gene symbols) in the combined cohort study and the mean number of non-benign DECIPHER SNVs in the overlapping genes (using gene symbols) and visualized as scatter plots with the top 25 GOMFs indicated.

Further, the top 50 most enriched GOMFs by significance (P value) were categorized into five major groups and colored accordingly in the scatter plots. Additionally, the total number of non-benign DECIPHER SNVs for the overlapping genes (using gene symbols) in these five major groups and chromatin remodeling complexes from the curated list of chromatin regulators were visualized as a bar chart (GOMF chromatin gene sets and chromatin regulatory complexes were merged into one group).

The GOMFs gene sets were obtained from MSigDB v7.5.1 (GOMF v7.5.1; https://www.gsea-msigdb.org/gsea/msigdb/). The ARID2, BCL7A/C and BICRAL BAF genes were added to the chromatin binding GOMF gene set.

Benign and likely benign SNVs in DECIPHER were excluded to create the set of non-benign DECIPHER SNVs. The development disorder DECIPHER gene list was based on DDG2P genes on DECIPHER (accessed on 15 May 22).

NDD recurrence in chromatin regulatory complexes, epigenetic modifiers and synaptic genes

Queried chromatin remodeling gene lists (Supplementary Table 2) were used for all downstream analysis in Fig. 1/Extended Data Fig. 1.

The total number of de novo missense and PTVs in the combined cohorts ASD + DD study (SSC/ASC, SPARK, and DDD) across a curated list of chromatin regulators and EpiFactor complexes were visualized as bar charts. The total number of de novo missense and PTVs in DD (DDD) and ASD (SSC/ASC and SPARK) across EpiFactor complexes were visualized separately as bar charts. The total number of de novo missense and PTVs in ASD + DD for every gene was visualized as a scatter plot with BAF genes indicated in red. The mean number of de novo missense and PTVs in ASD + DD (SSC/ASC, SPARK, and DDD) across EpiFactor complexes were visualized as a bar chart. Protein lengths were obtained from the top reviewed UniProtKB accession for each gene. The total de novo missense and PTVs in ASD + DD (SSC/ASC, SPARK, and DDD) for each EpiFactor complex was divided by the total protein length of each EpiFactor complex to obtain protein length-normalized NDD de novo mutational recurrence (that is average number of de novo missense and PTVs per residue in each EpiFactor complex). The protein length-normalized de novo mutational recurrence for EpiFactor complexes were visualized as a bar chart.

Benign and likely benign SNVs in DECIPHER were excluded to create the set of non-benign DECIPHER SNVs. The mean number of non-benign DECIPHER SNVs and de novo missense and PTVs in ASD + DD across all EpiFactor complex genes, mSWI/SNF genes and SynGO synaptic genes were visualized as bar charts. The total number of non-benign DECIPHER SNVs across a curated list of chromatin regulators were visualized as a bar chart.

All bar charts were created using matplotlib[130], and mSWI/SNF and cBAF, PBAF and ncBAF gene sets are indicated in red. Ensembl gene IDs (ENSG IDs) were used to overlap genes, merge datasets, and calculate

the total or mean number of de novo missense and PTVs in ASD + DD and non-benign DECIPHER SNVs for gene sets in the list of curated chromatin regulators and EpiFactor complexes (Supplementary Table 2).

Structure figures

The mapping of unique SNV and short in-frame insertion/deletion mutations was visualized using PyMol (v2.4.0) (ref. [133]). The structural models used for this study were the following: Recombinant cBAF structure bound to nucleosome (PDB: 6LTJ), Endogenous cBAF structure bound to nucleosome (PDBDEV: PDBDEV_00000056), PBAF complex bound to nucleosome (7VDV), SNF2h (5X0Y), yeast SWI/SNF (6UXW), ARID1A-ARID (1RYU), DPF2-PHD (5B79), SMARCE1-HMG (7CYU) and SMARCB1-WH (6LTJ). Domain annotations were obtained from the PFAM and the literature, and manually curated (Supplementary Table 3).

Conservation analysis

Conservation analysis was performed for the recombinant cBAF structure (PDB: 6LTJ; SMARCA4, chain I and SMARCB1, chain M), and the ARID1A-ARID (1RYU), DPF2-PHD (PDB: 5B79), and SMARCE-HMG (PDB: 7CYU) domains using the ConSurf Server (https://consurf.tau.ac.il/)[134]. Briefly, Protein Data Bank (PDB) IDs were selected and run through ConSurf analysis using standard parameters (HMMR search algorithm, UNIREF-90 protein database, automatic homolog selection and MAFFT multiple sequence alignment method). Once completed, amended PDB files color coded by conservation were downloaded and instructions to 'create high resolution figures' were followed as instructed by the ConSurf server.

Pairwise alignment

Multiple sequence alignments of the SMARCA4-ATPase, SMARCB1-CTD, ARID1A-ARID, SMARCB1-WH, DPF2-PHD and SMARCE1-HMG domains with their respective homologous proteins were performed using Geneious Prime (v2021.2.2) using standard parameters.

General

Unless otherwise noted, mutational counts, bar plots, heatmaps and pie charts throughout were made using a combination of R (v4.1.1), GraphPad Prism (v9.2.0) and matplotlib (v3.3.1), and seaborn.

ConSurf mutational analysis

Full-length FASTA sequences of the UniProt canonical transcript for all mSWI/SNF genes were uploaded to the ConSurf server with default parameters to obtain predicted conservation scores. The number of missense and in-frame indel NDD mutations by gene and position and the predicted ConSurf conservation score (negative-transformed so that higher scores indicate more conserved residues) were visualized as a scatter plot. All mSWI/SNF genes were used for this analysis.

NDD domain mutation analysis

The proportion of NDD mutations from the compiled list (missense, in-frame indels, frameshift and nonsense mutations) were summed for each gene, domain and inter-domain regions (Supplementary Table 3). The proportion of NDD mutations within domains (intradomain) and between domains (interdomain) were visualized as a stacked bar plot. Domains were defined by PFAM, UniProtKB, manual curation and resolved structures.

NDD disorder analysis

The proportion of NDD mutations from the compiled list (missense, in-frame indels, frameshift and nonsense mutations) falling within disordered (defined by MobiDB-lite; Supplementary Table 3) and structured regions were visualized as a stacked bar chart for individual BAF genes and BAF genes as a whole collection.

PolyPhen mutational analysis

The PolyPhen HumVar[92] model was used to predict the severity of each missense mutation in the list of compiled NDD mutations. The number of NDD missense mutations for each intradomain (within-domain) and interdomain (between-domain) region was divided by the lengths of these regions to calculate the average number of NDD missense mutations per residue for each interdomain or intradomain region. The PolyPhen HumVar predicted severity scores for each residue in each interdomain and intradomain were summed and divided by the length of each region to calculate the average PolyPhen HumVar predicted severity score for each inter-domain and intra-domain region. The average predicted PolyPhen HumVar predicted severity score and average number of NDD missense mutations were visualized as a scatter plot with interdomain and intradomain status indicated by color. All BAF genes were used for this analysis.

2D schematics

The distribution of gnomAD (v3) missense SNPs were visualized as a kernel density estimate plot using the seaborn kdeplot with default parameters. The gnomAD (v3) missense mutations for SMARCA2, SMARCA4, ARID1A, ARID1B, SMARCB1, SMARCE1 and DPF2 were used to compute the missense recurrence by position across the length of each protein, which was used as input into the kernel density estimate analysis. The NDD compiled list of mutations (missense, in-frame indels, frameshift and nonsense mutations) for the aforementioned genes were visualized using the St. Jude PeCan Protein Paint software with default settings (https://proteinpaint.stjude.org/). Special care was taken to map the mutations on the canonical UniProt isoform (Supplementary Table 3). Domains using the annotations compiled from PFAM, InterPro and the literature, and manually curated based on the AlphaFold EMBL-EBI structural predictions. ConSurf conservation scores were visualized as horizontal bars using the ConSurf provided 'COLOR' column with an aggregation of scores (1, 2 or 3, cyan; 4, 5 or 6, white; 7, 8 or 9, violet). The coverage of the two available recombinant (PDB: 6LTJ) and endogenous nucleosome-bound cBAF structures were visualized as horizontal bars (recombinant coverage in orange, endogenous coverage in red and dual coverage in brown).

Missense DNA and protein changes

The frequencies of DNA point substitutions (all SNVs) and protein amino acid substitutions (top 20) in the compiled NDD mutation dataset (missense only) were visualized as bar plots. Additionally, the amino acid substitutions for the missense subset of mutations in the compiled NDD mutation dataset was visualized as Sankey Diagram using Google Charts. Additionally, these amino acid substitutions were aggregated into functional changes (negative, positive, polar, nonpolar and miscellaneous) and visualized as proportions in stacked bar charts.

Mappability of NDD mutations

The proportion of NDD mutations in the compiled NDD mutation dataset (missense, in-frame indels, frameshift and nonsense mutations) mappable across the endogenous and recombinant (PDB: 6LTJ) were visualized as a group bar plot (Supplementary Table 3).

NDD versus cancer overlap analysis

The recurrence of every unique gene-mutation combination for missense and in-frame indel mutations from the NDD compiled dataset and the cBioPortal (accessed June 2022) cancer dataset was computed and visualized as a pie chart or tables.

NDD versus cancer NESs and comparative analyses

The missense and in-frame indel mutations from the compiled NDD mutation dataset and the cBioPortal cancer dataset were used to compute the NDD and cancer mutation recurrence by position across each BAF gene. This recurrence was scaled between 0 and 1 using the

MinMaxScaler preprocessing function in scikit-learn. The rescaled mutation recurrence for cancer was subtracted from the rescaled mutation recurrence for NDD to compute the NDD-Cancer normalized enrichment scores (NESs). Specifically, cancer NESs were calculated using a four-step process. First, paralogs were pairwise aligned to the primary paralog, and mutations on conserved residues were remapped from the secondary to the primary paralogs. Second, the mutational recurrence by residue position of NDD- and cancer-associated missense and in-frame indel mutations were calculated across all mSWI/SNF subunits and averaged over a window size of 21 aa centered at each residue (10 amino acids on each side). Third, these smoothed averages were scaled to a range between 0 (no recurrence) and 1 (highest recurrence) to generate the local recurrence of NDD- and cancer-associated missense and in-frame indel mutations. Fourth, the local recurrence maps across all mSWI/SNF for NDD- and cancer-associated mutations were subtracted (NDD-cancer) to form the NDD-cancer NES on a range bounded by −1 (maximally enriched in cancer) and 1 (maximally enriched in NDD). NDD- and cancer-associated missense and in-frame mutations were derived as described in (Fig. 5a). These local and NESs were visualized across the specific paralogs in the recombinant cBAF structure (PDB ID 6LTJ) as various colored heatmaps (local NDD recurrence scaled in green, local cancer recurrence scaled in red, NDD-Cancer NESs in blue-white-red: blue = enriched in cancer, red = enriched in NDD) and across specific paralogs indicated in the Circos plot as a purple-orange histogram (purple, enriched in cancer; orange, enriched in NDD). The local enrichment scores for NDD (green) and cancer (red) were visualized as histograms in the outer bands of the Circos plot. Previously published nucleosome-bound cBAF cross-linking mass spectrometry data were combined and visualized as inner links on the Circos plot, where link thickness is proportional to the frequency of cross-links (the maximum frequency of cross-links is capped at 10 units). The Circos plot was made using the Circos software[135].

Rolling averages of cancer and NDD mutational recurrence (missense and in-frame indels only) were calculated for BAF genes and visualized as a scatter plot with a regression line using the seaborn[136] regplot function.

NDD functional mutation analysis
Specific NDD residues predicted (by structural analysis) to disrupt buried residues (altering cavities), buried charged residues and hydrogen-bonds, BAF subunit or BAF module interaction, and BAF domain interaction were visualized in PyMol, with the disruptive NDD mutations indicated in red and putative interacting/proximal residues in blue or purple. Additionally, Missense 3D webserver with recombinant NCP-bound cBAF complex as input was used to assign functional consequences of some of these disruptive NDD mutations.

NDD human versus yeast analysis
Select NDD residues in the integrated dataset were mapped to the recombinant NCP-bound cBAF complex (PDB: 6LTJ), yeast Swi/Snf (PDB:6UXW) and Snf2-nucleosome structures (PDB:5X0Y, 5X0X) were used to show that seemingly exposed residues on the cBAF structure are in fact buried by the brace helices in SMARCA2/A4 and that certain side-chain orientations in cBAF structure have different orientations in the yeast structures. SMARCA2/4 variant residues were mapped onto additional yeast Snf2-nucleosome structures (PDB:5Z3O, 5Z3U) to explore the open (ADP-bound) and closed (ADP-BeFx-bound) ATPase states and emphasize ATP and DNA interacting residues of the ATPase domain.

Statistics and reproducibility
A hypergeometric test was used to determine the enrichment of genes of interest in a given gene set representing a specific biological process, molecular function, pathway or meaningful biological collection of genes. This analysis is more thoroughly described under NDD Gene Set Enrichment Analysis. OLS regression analysis was carried out using the default parameters in the seaborn regplot function.

No statistical method was used to predetermine sample size. Samples sizes for the hypergeometric test were determined using the standard procedure for GO, enrichment, or overrepresentation analysis.

Known duplicate samples or potentially duplicate samples from manual curation were excluded from analysis. Criteria for exclusion are thoroughly described under Curating mSWI/SNF gene NDD-associated variants. No other data were excluded from the analyses from variants collected from the aforementioned public or private databases. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Reporting summary
Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability
Public and private data can be accessed through their respective portals. Private data will require prior authorization. Data can be cleaned and normalized using any standard or well-established procedure for variant analysis or the procedures described in this paper, including referenced papers or procedures. The integrated, curated and de-duplicated data (to the best of our ability) are available in Supplementary Table 1. No additional data or intermediate results will be available upon request given the high manual burden to verify access to a variety of private portals, repositories and patients.

Code availability
Variants were processed using well-established procedures described in the referenced papers. Datasets from diverse sources were integrated using a combination of code (to automate certain steps) and manual curation. Thus, the standalone code is not sufficient to regenerate the integrated dataset. Therefore, this code and intermediate results from dataset integration and curation is not available upon request. The code used for analysis and to generate figures is available under Creative Commons license through Zenodo at https://doi.org/10.5281/zenodo.8008632. Analyses were executed in Python (v3.7), R (v4.1.1), GraphPad Prism (v92.2), matplotlib(v3.3.1), circos (v0.69-9) and seaborn (v0.11.1).

PyMOL v2.4.0 was used to visualize structures. The Consurf online server was used for conservation analysis. Geneious Prime v2021.2.2 was used for multiple sequence alignmentss. The PolyPhen2 online server using the HumVar model was used to predict the severity/pathogenicity of the compiled NDD mutations. Unless otherwise noted, mutational counts, bar plots, pie charts, and Venn diagrams throughout were made using a combination of Python (v3.7), R (v4.1.1), GraphPad Prism (v92.2), matplotlib(v3.3.1) and seaborn (v0.11.1). The lollipop portion of the 2D schematics were created using the St. Jude PeCan Protein Paint software. Missense substitutions were visualized as a Sankey diagram using Google Charts. The Circos plot was made using the Circos software (v0.69-9). Missense substitutions were visualized as a Sankey diagram using Google Charts. The code used to process and visualize the data are available under the MIT license at Zenodo at https://doi.org/10.5281/zenodo.8008632.

References
116. Hargreaves, D. C. & Crabtree, G. R. ATP-dependent chromatin remodeling: genetics, genomics and mechanisms. *Cell Res.* **21**, 396–420 (2011).
117. Sokpor, G., Castro-Hernandez, R., Rosenbusch, J., Staiger, J. F. & Tuoc, T. ATP-dependent chromatin remodeling during cortical neurogenesis. *Front. Neurosci.* **12**, 226 (2018).

118. Li, Y. et al. The emerging role of ISWI chromatin remodeling complexes in cancer. *J. Exp. Clin. Cancer Res.* **40**, 346 (2021).

119. Torrado, M. et al. Refinement of the subunit interaction network within the nucleosome remodelling and deacetylase (NuRD) complex. *FEBS J.* **284**, 4216–4232 (2017).

120. Sardiu, M. E. et al. Conserved abundance and topological features in chromatin-remodeling protein interaction networks. *EMBO Rep.* **16**, 116–126 (2015).

121. Giaimo, B. D., Ferrante, F., Herchenröther, A., Hake, S. B. & Borggrefe, T. The histone variant H2A.Z in gene regulation. *Epigenetetics Chromatin.* **12**, 37 (2019).

122. Fröb, F. & Wegner, M. The role of chromatin remodeling complexes in Schwann cell development. *Glia* **68**, 1596–1603 (2020).

123. Willhoft, O. & Wigley, D. B. INO80 and SWR1 complexes: Rhe non-identical twins of chromatin remodelling. *Curr. Opin. Struc. Biol.* **61**, 50–58 (2020).

124. Conaway, R. C. & Conaway, J. W. The INO80 chromatin remodeling complex in transcription, replication and repair. *Trends Biochem. Sci.* **34**, 71–77 (2009).

125. Shah, S. G. et al. HISTome2: a database of histone proteins, modifiers for multiple organisms and epidrugs. *Epigenetetics Chromatin* **13**, 31 (2020).

126. Khare, S. P. et al. HIstome—A relational knowledgebase of human histone proteins and histone modifying enzymes. *Nucleic Acids Res.* **40**, D337–D342 (2012).

127. Croce, L. D. & Helin, K. Transcriptional regulation by Polycomb group proteins. *Nat. Struct. Mol. Biol.* **20**, 1147–1155 (2013).

128. Greenberg, M. V. C. & Bourc'his, D. The diverse roles of DNA methylation in mammalian development and disease. *Nat. Rev. Mol. Cell Biol.* **20**, 590–607 (2019).

129. Fu, J. M. et al. Rare coding variation provides insight into the genetic architecture and phenotypic context of autism. *Nat. Genet.* **54**, 1320–1331 (2022).

130. Hunter, J. D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).

131. McLaren, W. et al. The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 (2016).

132. Wall, L., Christiansen, T., & Orwant, J. *Programming perl* (O'Reilly Media, 2000).

133. Schrödinger, L., & DeLano, W. *PyMOL*. http://www.pymol.org/pymol (2020).

134. Ashkenazy, H. et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* **44**, W344–W350 (2016).

135. Krzywinski, M. et al. Circos: An information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645 (2009).

136. Waskom, M. seaborn: statistical data visualization. *J. Open Source Softw.* **6**, 3021 (2021).

Acknowledgements

Author contributions

A.M.V. and C.K. conceived of and directed the study. A.S. performed all computational and statistical analyses. F.K.S., J.F. and M.T. analyzed and curated the SFARI and DDD datasets used in this analysis. P.J.v.d.S. curated and presented newly reported NDD-associated mutations. S.A.S.V. and G.W.E.S. curated and contributed novel human genetic sequencing data and edited the manuscript. C.K. and A.M.V. wrote the manuscript and all authors critically reviewed and edited the manuscript.

Competing interests

C.K. is the scientific founder, scientific advisor to the Board of Directors, scientific advisory board member, shareholder and consultant for Foghorn Therapeutics. C.K. is also a member of the scientific advisory board and is a shareholder of Nested Therapeutics, Nereid Therapeutics and Accent Therapeutics, serves on the scientific advisory board for Fibrogen and serves as a consultant for Google Ventures and Cell Signaling Technologies. C.K. and A.M.V. hold patents in the field of mSWI/SNF complex targeting therapeutics. S.A.S.V. is a member of the scientific advisory board at Ambry Genetics, for which no compensation is received. The other authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41588-023-01451-6.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41588-023-01451-6.

Correspondence and requests for materials should be addressed to Cigall Kadoch.

Peer review information *Nature Genetics* thanks the anonymous reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.
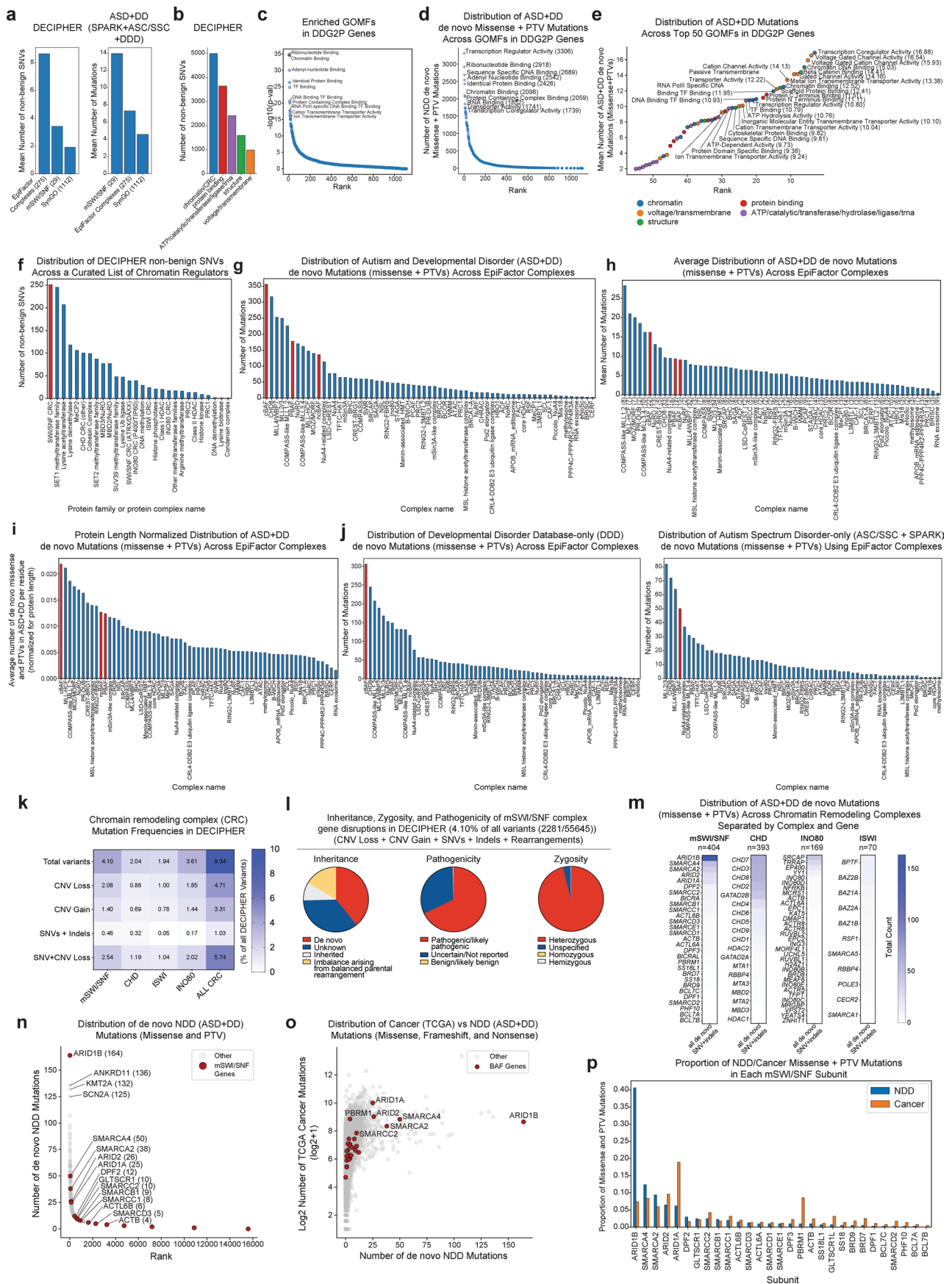
Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | SWI/SNF complex genes are among the most frequently mutated genes in human NDD. a, Bar charts depicting mean number of non-benign SNVs in DECIPHER and ASD+DD across gene sets indicated. **b**, Bar graph summarizing the number of non-benign DECIPHER SNVs across top 5 categories from Fig. 1a. **c**–**e**, Rank plots depicting GOMF gene sets in (c) DDG2P, (d) ranked by total number of ASD+DD de novo missense variants, (e) ranked by mutation frequency, top 50 GOMFs. **f**–**j**, Bar charts showing distribution of variants across sets indicated in each title. mSWI/SNF or cBAF, PBAF, and ncBAF are highlighted in red. **k**, Heatmap summarizing DECIPHER database mutational frequency for each chromatin remodeling complex separated by variant type (all variants, copy number variants (CNV), and SNVs/indels). **l**, Pie charts showing inheritance, pathogenicity, and zygosity breakdown of all mSWI/SNF complex variants from DECIPHER. **m**, Heatmaps depicting the mutational frequency of chromatin remodeling genes in SWI/SNF, CHD, ISWI, and INO80 complex family classes in the ASD+DD dataset. Total number of SNV and indel variants per protein complex family indicated. **n**, Scatterplot of the total number of de novo missense and PTVs in ASD+DD for all genes ranked by the mutational burden of each gene. mSWI/SNF genes are shown in red. **o**, Scatterplot of the log normalized total number of cancer missense, frameshift, and nonsense mutations in the TCGA MC3 PanCancer dataset versus the total number of NDD de novo missense and PTVs in ASD+DD datasets. mSWI/SNF genes shown in red. **p**, Grouped bar graph of the proportion of NDD (blue) and cancer (orange) missense and PTV mutations across all mSWI/SNF genes sorted by decreasing NDD mutational proportion.
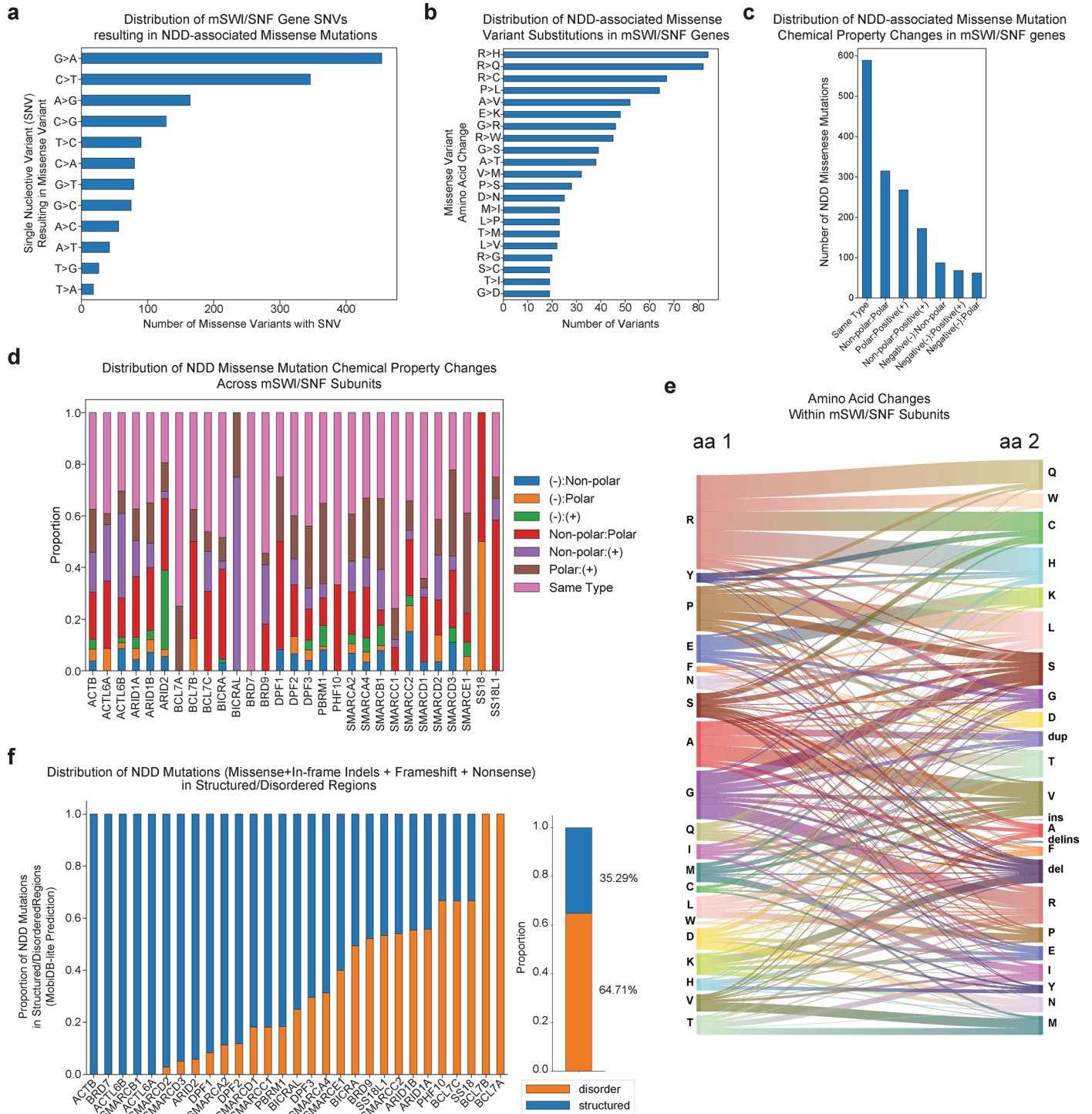
Extended Data Fig. 2 | Characteristics of NDD-associated single-residue amino acid perturbations in mSWI/SNF components. a, Distribution of single-nucleotide variants (SNVs) found in NDD-associated missense mutations of mSWI/SNF family genes (Supplementary Table 1) in the integrated dataset (n=2539). **b**, Horizontal bar graphs of the top 20 amino acid missense substitutions in the integrated dataset (Supplementary Table 1). **c**, Bar chart characterizing amino acid chemical property changes upon missense mutation for NDD-associated variants in the integrated dataset. **d**, Stacked bar graphs

of the distribution of amino acid substitution chemical property changes in NDD-associated missense mutations in the integrated dataset. **e**, Sankey diagram of the distribution of NDD-associated missense substitutions in the integrated dataset. Ribbon thickness represents frequency of substitutions in the integrated dataset. **f**, Stacked bar chart summarizing percentage of NDD-associated missense and in-frame indel mutations in the integrated dataset falling within intrinsically disordered (defined by MobiDB-lite) or structured regions for (left) each mSWI/SNF subunit and (right) all mSWI/SNF subunits combined.
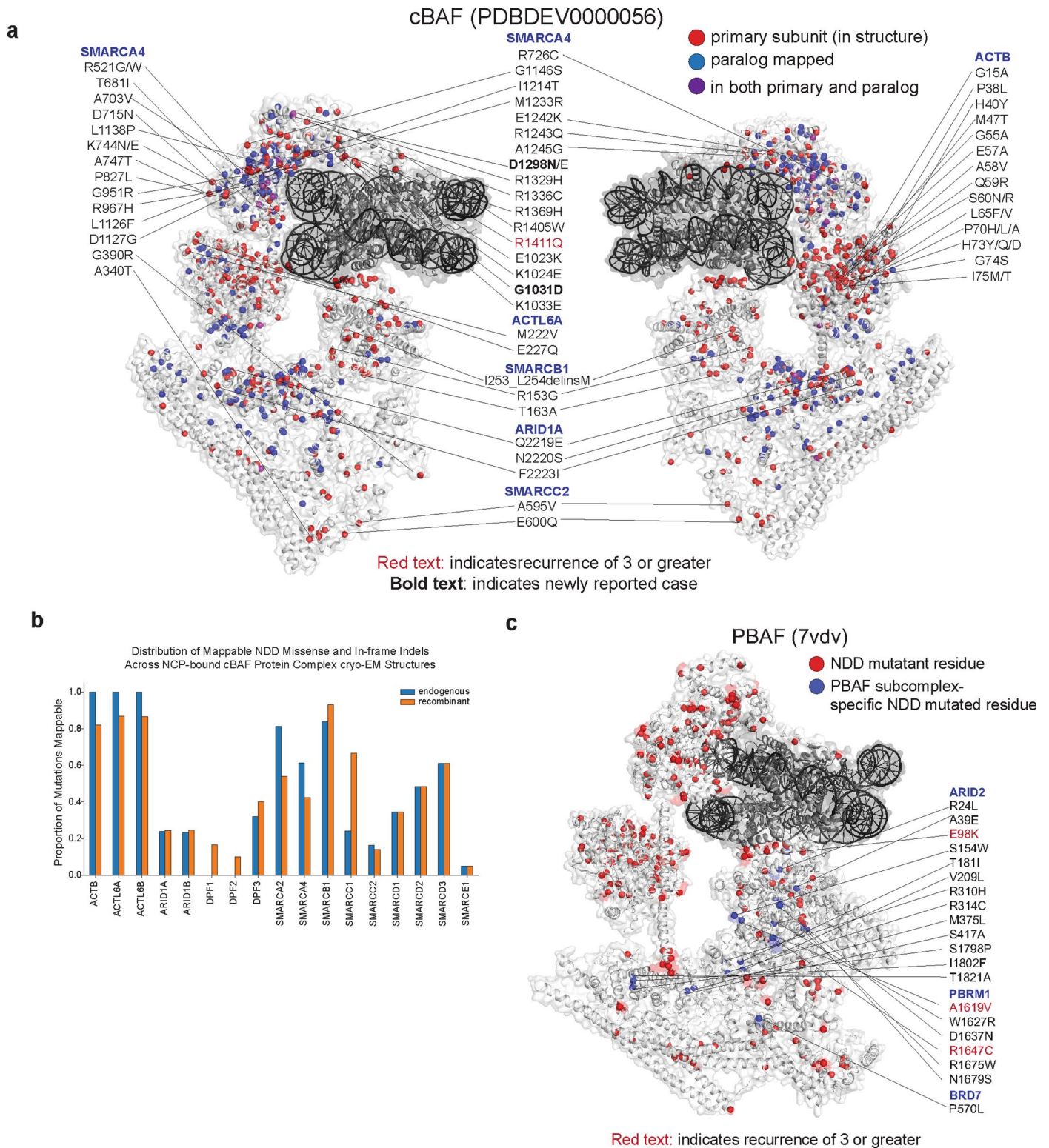
Extended Data Fig. 3 | NDD-associated missense variants mapped on cBAF and PBAF 3D structures. a, NDD-associated missense and inframe indel variants mapped on to the 3D structure of the endogenous human cBAF complex (PDBDEV_00000056). Red spheres represent NDD-associated variants in the subunit indicated, blue spheres represent those mapped from the paralog subunit, and residues in purple represent NDD-variants mapped in both primary subunit present on cBAF structure and paralog subunit. Variants that map exclusively on endogenous complex are indicated. Recurrent variants (n>3) are emphasized in red. **b**, Bar chart indicating proportion of NDD-associated missense and in-frame indel mutations in the integrated dataset mappable to current mSWI/SNF complex structures separated by subunits. **c**, NDD-associated missense and inframe indel variants mapped on to the 3D structure of the PBAF complex (PDB 7VDV). Red and blue spheres represent NDD-associated variants in the subunit indicated. Blue spheres and annotations emphasize PBAF subcomplex specific variants mapped.
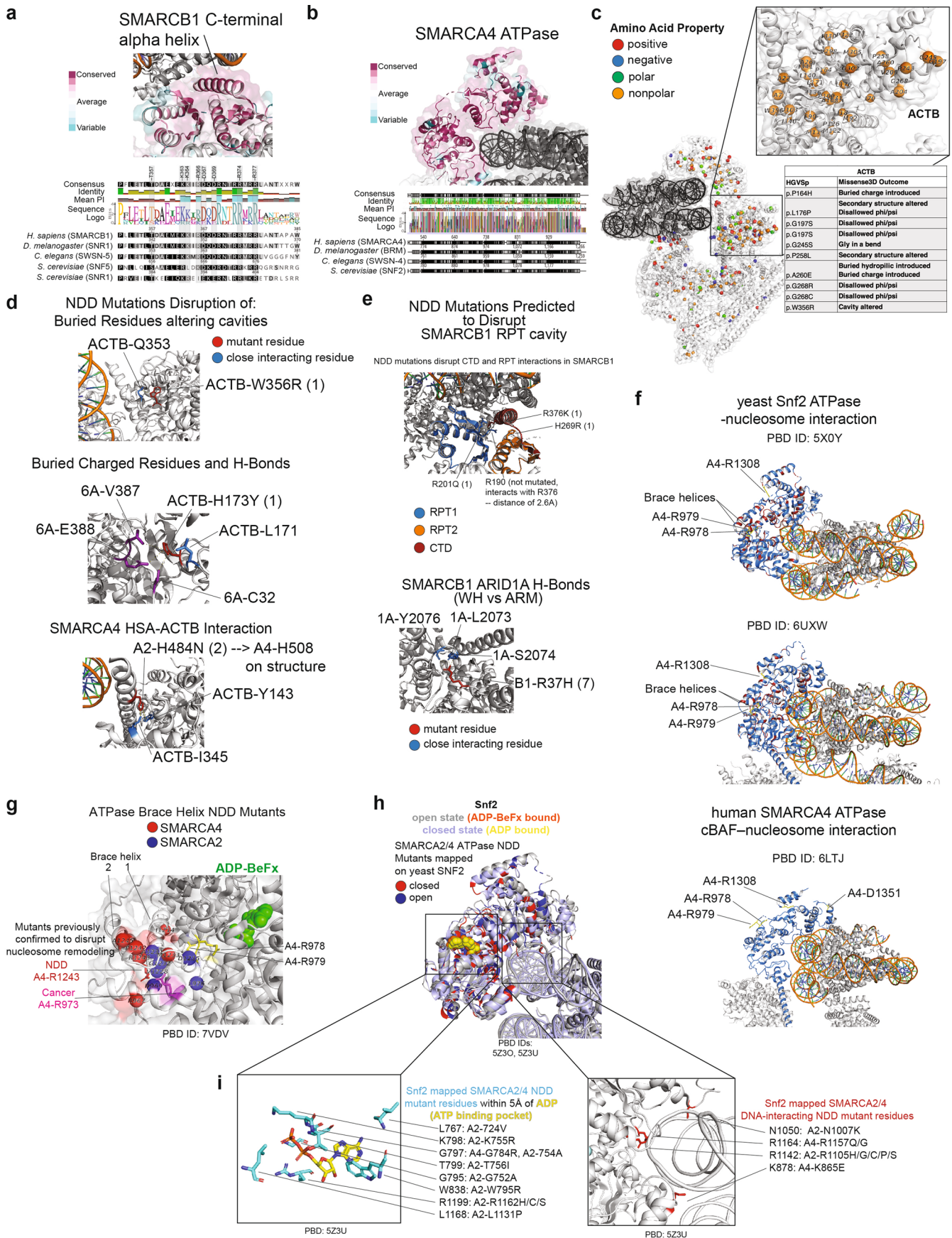
Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Structural dissection of mSWI/SNF subunit mutations across the ARP, Core, and ATPase modules. a, b, (a) SMARCB1-C terminal alpha helix and (b) SMARCA4-ATPase domain (top) ConSurf conservation mapping and (bottom) multiple sequence alignment using D. melanogaster, C. elegans, and S. cerevisiae homologs. **c**, NDD-associated missense and in-frame indel variants mapped onto the 3D structure of the cBAF complex (PDB:6LTJ) color coded by residue chemical characteristics: red: positive charge, blue: negative charge, green: polar, orange: nonpolar. Nonpolar residues of the ACTB (Arp module) and Table of nonpolar mutations predicted to structurally disrupt ACTB are shown. **d**, ACTB NDD mutations may alter internal hydrophobic cavities, interfaces with ACTL6A/B, and interfaces with SMARCA2/A4-HSA. Mutant residues shown in red and putative proximal/interacting residues shown in blue/purple. **e**, SMARCB1-RPT and WH domain NDD mutations predicted to disrupt internal cavity integrity, and hydrogen bonding to interacting ARID1A main chain carbonyls, respectively. Top, selected NDD-associated SMARCB1 missense mutations are labeled, and major domains of SMARCB1 are colored, including RPT1 (blue), RPT2 (orange), and CTD (red). Bottom, mutant residue shown in red and putative proximal/interacting residues shown in blue. **f**, Mapping of conserved SMARCA2/4 NDD mutant residues (red) on the yeast Snf2 ATPase domain (5X0Y and 6UXW) compared to the recombinant cBAF SMARCA4 ATPase (6LTJ). Brace helices (indicated in yeast structures) are not resolved in human cBAF structure, but demonstrate that certain residues, emphasized in yellow, are buried by the SMARCA2/4 brace helices, rather than exposed. **g**, Mapping of SMARCA2/4 brace helix NDD variants onto the closed state of the SMARCA4 ATPase domain using the PBAF structure (7VDV). NDD variants clustered in brace helices are predicted to disrupt nucleosome remodeling activity as has been shown with R1243 and R973 NDD and cancer-associated mutations indicated in panel[97]. **h**, Mapping of SMARCA2/4 NDD mutant residues on the Snf2 ATPase open (gray) and closed (pale cyan) states (PDB IDs: 5Z3O, 5Z3U). NDD residues colored blue in open state and red in closed state. **i**, SMARCA2/4 NDD mutant residues (left) within 5Å of the ADP-BeFx and (right) interacting with nucleosomal DNA mapped onto the closed yeast Snf2 ATPase structure (5Z3U).
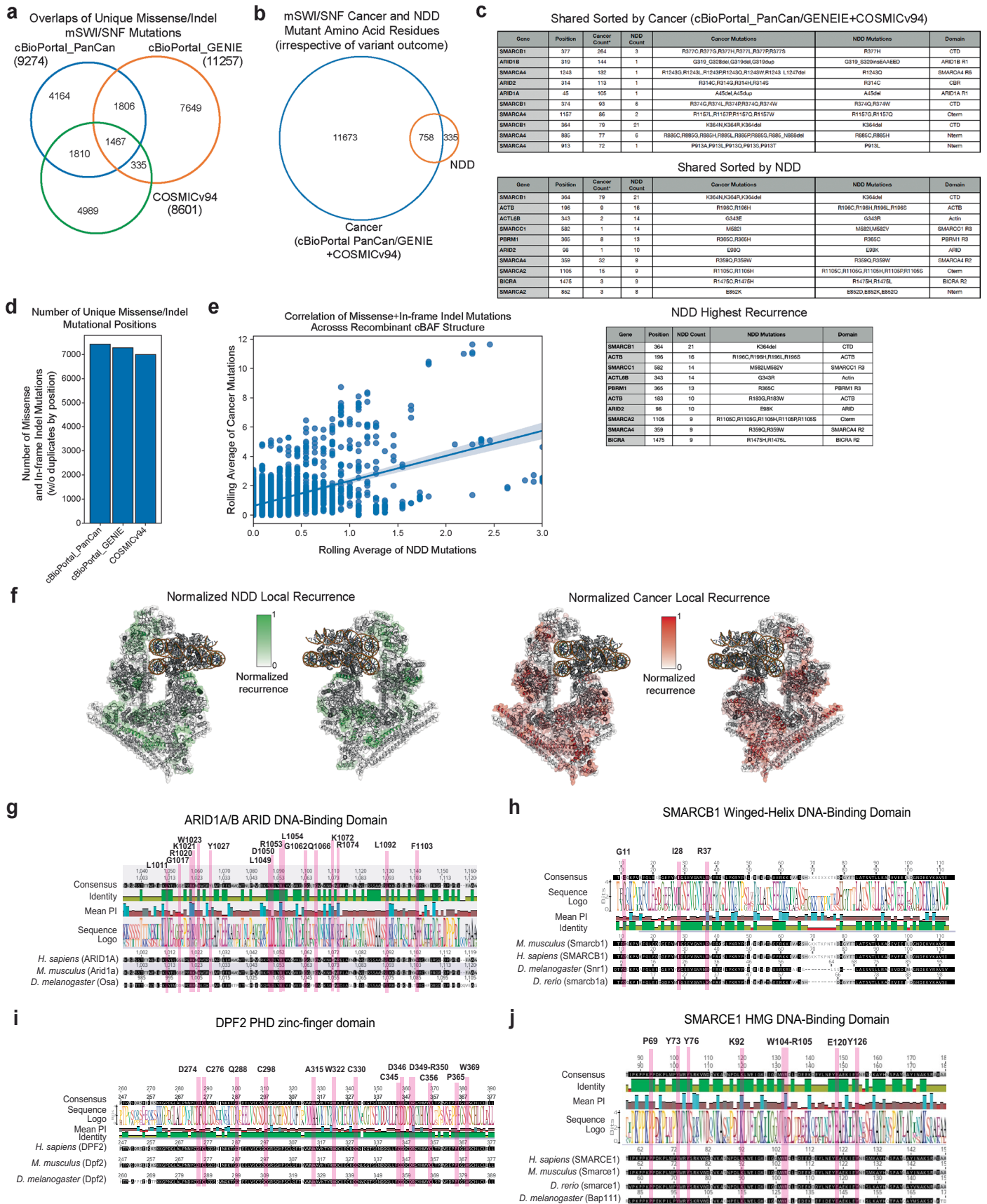
Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Perturbed subunit positions shared between cancer and NDD highlight ATPase, nucleosome binding regions, and Arp module.
a, Venn diagram overlapping unique cancer missense and inframeshift variants identified from cBioPortal_PanCan, cBioPortal_GENIE and COSMICv94 cancer genetics datasets. **b**, Venn diagram overlapping unique cancer and NDD (Supplementary Table 1) missense and inframe variants by amino acid position regardless of mutation consequence. NDD mutations derived from Supplementary Table 1, cancer mutations derived by combining cBioPortal_PanCan, cBioPortal_GENIE and COSMICv94 datasets. **c**, Top ten most recurrent mutant residue amino acid positions shared between Cancer and NDD sorted by frequency in each disease type. Highest recurrence of NDD mutations also included. NDD- and cancer-associated mutations were derived as described in (b). **d**, Bar plot showing the total number of unique missense/indel mSWI/SNF mutations across the following cancer datasets: cBioPortal_PanCan, cBioPortal_GENIE, COSMICv94. **e**, Correlation of missense and inframeshift mutations shared between cancer (cBioPortal_PanCan only) and NDD across recombinant cBAF structure. Briefly, NDD- and cancer-associated missense

and in-frame indel mutations were remapped onto the primary paralogs of the recombinant cBAF (PDB ID: 6LTJ) structure. A rolling average with a window size of 11aa centered on each residue (5aa on each side) of mutation recurrence by residue position for NDD and cancer was used for the scatterplot and correlation calculation. NDD- and cancer-associated mutations were derived from Supplementary Table 1 (NDD) and cBioPortal_PanCan datasets. The translucent bands around the regression line represent the 95% confidence interval estimated using a bootstrap for 100 iterations. **f**, Heatmap representation of scaled local enrichment of NDD- and cancer-associated missense and in-frame indel mutational burden of (left, in green) NDD and (right, in red) cancer reflected on the 3D structure of the human cBAF complex (PDB: 6LTJ). Local enrichment scores were computed as described in (Fig. 5e). NDD- and cancer-associated mutations were derived as described in (Fig. 5e). **g–j**, Multiple sequence alignment of (g) ARID1A-ARID domain, (h) SMARCB1-WH domain, (i) DPF2-PHD domain, and (j) SMARCE1-HMG domain, with variety of related homologs (including M. musculis, D. rerio, D. melanogaster, C. elegans, and S. cerevisiae, where possible).

nature portfolio

Corresponding author(s):	Cigall Kadoch, Ph.D.
Last updated by author(s):	06/05/2023

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Data collection was done as described in the the Methods section of the manuscript and copied in the "Data" section below. Novel mSWI/SNF-NDD variants were identified as described in the "Novel Variant Collection" in the Methods of the manuscript, and summarized briefly in the "Human Research Participants" section in this document.
Data analysis	PyMOL v2.4.0 was used to visualize structures. The Consurf online server was used for conservation analysis. Geneious Prime v2021.2.2 was used for MSAs. The PolyPhen2 online server using the HumVar model was used to predict the severity/pathogenicity of the compiled NDD mutations. Unless otherwise noted, mutational counts, bar plots, heatmaps, and pie charts throughout were made using a combination of R (v4.1.1), GraphPad Prism (v9.2.0), and matplotlib (v3.3.1). The lollipop portion of the 2D schematics were created using the St. Jude PeCan Protein Paint software. Missense substitutions were visualized as a Sankey diagram using Google Charts. The circos plot was made using the circos software. The code used to process and visualize the data are available on request.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Mutational Datasets: Open-access mutations publicly available on the DECIPHER database (https://www.deciphergenomics.org/; accessed June 22, 2022) (Firth et al., AJHG 2009) were used for broader chromatin gene analysis (Fig. 1c-d, S1k-l). The queried chromatin remodeling complex gene list (SWI/SNF, CHD, INO80, and ISWI) was manually curated from a literature review detailed below (Supplementary Table 2).

Chromatin Regulatory Gene Set (for Table S2): Chromatin remodeling complex (CRC) gene lists were curated from a variety of sources including HGNC gene groups SWI/SNF, INO80 (https://www.genenames.org/data/genegroup/#!/), as well as a literature review of the following: all CRCs– Sokpor et al., Front Neuro (2018), Hargreaves & Crabtree, Cell Res, (2011), mSWI/SNF– Mashtalir et al., Cell (2018), ISWI– Li et al, J Exp Clin Cancer Res (2021), CHD: Torrado et al., FEBS J 2017, and INO80: Sardiu et al, Sci Rep (2015), Giaimo et al, Epigenetics & Chromatin (2019), Frob & Wegner, Glia (2019), Willhoft & Wigley, Curr Opinion in Str Biol (2020), Feng et al, Cell Research (2018), Conaway & Conaway, Trends in Biochemical Sciences (2009). Histone modifier gene list was gathered from HISTome2 (Shah et al., Epigenetics & Chromatin (2020), Khare et al., Nucleic Acids Research (2012); http://www.actrec.gov.in/histome2/). Polycomb repressive complex genes were informed by Di Croce & Helin, Nat Struct Mol Biol (2013). DNA methylation regulatory genes were informed by Greenberg & Debroah Bourc'his, Nat Rev Mol Cell Biol (2019). Additional chromatin regulatory complexes were obtained from EpiFactor (Medvedeva et al., Database (2015); https://epifactors.autosome.org/protein_complexes). The full set of cBAF, PBAF, and ncBAF genes were included in the EpiFactor complexes if absent.

Curating mSWI/SNF gene NDD-associated variants: The set of rare inherited and de novo variants included data from three cohorts of individuals with autism spectrum disorders or other developmental disorders: the Simons SSC/ASC, SPARK, and DDD cohorts. Details about merging and de-duplicating the data are described in Fu et al. Briefly, duplicated samples were identified and excluded by IBD and other metadata, and the filtered samples were merged to provide a single unified set of de-duplicated de novo variants in autism spectrum disorders and other developmental disorders. The recurrence of NDD de novo variants across BAF genes and several gene sets of interest, including a curated set of chromatin remodelers, epigenetic modifiers, and synaptic genes were visualized with scatter plots and bar charts using matplotlib. The set of de novo variants and non-benign SNVs in DECIPHER were used for all summary calculations in figures 1 and S1, and for comparisons between the BAF genes, chromatin regulatory genes, epigentic modifier genes, and synaptic genes. The queried chromatin regulatory gene list was based on EpiFactor (https://epifactors.autosome.ru/genes; accessed September, 2 2021) (Medvedeva et al., Database 2015) updated to include all mSWI/SNF genes (Supplementary Table 2). The queried synaptic gene list was based on the SynGO gene list (https://www.syngoportal.org/ ; accessed September, 2 2021) (Koopmans et al., Neuron 2019). The development disorder DECIPHER gene list was based on the Development Disorder Genotype – Phenotype Database (DDG2P) genes on DECIPHER (accessed June 13, 2022).

A comprehensive list of single-nucleotide variant (SNV) and short in-frame shift mutations was compiled from an extensive literature review, the combined set of rare inherited and de novo variants from the Simons SSC/ASC, SPARK, DDD cohorts (the "combined cohort study"), the DECIPHER database of SNVs (https://www.deciphergenomics.org/), the merged set of de novo mutations from the DNM effort by McRae et al., NDD-associated ClinVar mutations (accessed 5/15/2021), NDD-associated variants from LOVD (LOVD v3.0 accessed 06/2022), and 85 previously unreported cases published in this study collected from the laboratories of Dr. Vergano and Dr. Santen.

First, the combined set of rare inherited and de novo variants were split into two: a set of rare inherited variants and a set of de novo variants. All rare inherited PTVs, in-frame indel variants, and de novo variants were included in the integrated dataset. Guided by the analysis in Fu et al., where missense variants with MPC scores (Missense badness, PolyPhen-2, and Constraint) of 1 or more were observed to confer moderate to strong levels of risk in developing autism, missense rare inherited variants with MPC scores >=1 were included in the integrated dataset. All other rare inherited variants from the combined cohort study were excluded. Then, samples were cross-referenced between the combined cohort study, DECIPHER database, and the DNM cohort of de novo mutations and identical variants from the same samples (using available sample IDs or aliases) were removed to de-deduplicate the data between these three cohorts / databases. Separately, a list of de novo variants in BAF genes across several other studies in the literature not covered previously by the cohorts used in DECIPHER and the combined cohort study (SSC/ASC, SPARK, DDD) were manually curated and de-duplicated to form the compiled set of mutations from the literature. Additionally, NDD-associated mutations from the LOVD database were compiled and filtered to include all PTV and in-frame indels and de novo/likely de novo missense variants. All benign/likely benign variants were excluded. The filtered set of LOVD variants and the manually curated variants from the literature were merged and de-duplicated based on sample IDs or aliases (if available) and study ID / reference (if sample IDs were not available). For shared variants between LOVD and the literature where it was not clear whether these variants were duplicates, only shared variants from the manually curated literature dataset were kept, effectively de-duplicating the data. Minimal overlap was assumed between the de-duplicated set of LOVD/literature variants and the de-duplicated set of SSC+ASC/SPARK/DDD/DECIPHER/DNM variants. These two sets were merged, followed by a round of manual curation to double check that as many duplicates or potential duplicates were removed during dataset integration. The set of 85 novel cases identified by Drs. Vergano and Santen were added to this merged dataset. In parallel, a curated set of ClinVar variants from samples with NDD-associated clinical features and unknown/likely pathogenic/pathogenic clinical significance was generated. Benign and likely benign ClinVar variants were excluded. Additionally, ClinVar variants submitted by GeneDx were excluded due to significant overlap with the comprehensive analysis of de novo mutations in NDD by Kaplanis et al. included in the DNM database of de novo mutations. Samples were de-duplicated between ClinVar and the LOVD/literature dataset using SCV codes wherever available. Finally, this de-duplicated ClinVar dataset was used to adjust the counts of the previously merged dataset of NDD-associated BAF mutations from the combined cohort study (SSC/ASC, SPARK, and DDD), DECIPHER SNVs, DNM, LOVD, and the literature. Note, it was difficult (and sometimes impossible) to track, match, and assign each filtered NDD-associated ClinVar SCV (submitted record for each variant) with the list of available sample IDs or aliases in the previously merged dataset. Thus, the total counts for each variant were adjusted to the total counts found in ClinVar (based on the number of submissions for each variant using SCV IDs) to eliminate the possibility of double counting if the ClinVar total count for a variant was more than the total count from the previously merged dataset. This procedure assumes submissions to ClinVar overlap entirely with the previously merged dataset, so it is possible the new merged dataset containing ClinVar variants might undercount some NDD-associated BAF variants. This integrated dataset was compared to gnomAD v3.1.2 to remove potential SNPs and other variants that occur frequently in a collection of healthy individuals. A more stringent MAF threshold of >=0.5% MAF was used to exclude potentially common variants in the integrated dataset. This final integrated dataset was manually checked once more to exclude potential duplicates and likely benign variants before freezing for all downstream analyses. A total of 2539 NDD-associated BAF variants are included in this dataset, including 85 novel cases, including 72 novel variants.

To standardize the data, all variants were remapped to the UniProt canonical BAF protein isoforms (see Supplementary Table S3) and duplicates, that could not be confirmed unique cases, were removed.

gnomAD variants of the general population were derived from the gnomAD v3 dataset (accessed 1/11/2021).

Cancer Datasets Cleaning and Compilation: PanCancer datasets from TCGA and cBioPortal were cleaned and compiled for all downstream analyses related to NDD vs Cancer comparisons.

The TCGA MC3 PanCancer dataset was used for NDD vs Cancer comparisons in FS1. Briefly, known SNPs were removed and BAF gene mutations were remapped to the canonical UniProt transcripts (see Supplementary Table S3). Missense, nonsense, and frameshift mutations were included, and all other mutations were excluded. This filtered set of mutations merged with the combined cohort study of NDD-associated mutations from the combined SSC/ASC, SPARK, and DDD cohorts. Total cancer missense, frameshift, and nonsense mutational recurrence was log normalized, compared to total de novo NDD-associated missense and PTV mutational recurrence for each gene, and visualized as a scatterplot using matplotlib with BAF genes indicated in red. The total proportion of NDD and Cancer missense and PTV mutations across the BAF genes were visualized as a grouped bar chart using matplotlib.

Mutations across BAF genes from the curated set of non-redundant studies in cBioPortal were compiled and filtered for NDD vs Cancer comparative analyses across the BAF genes. Briefly, the BAF mutations were remapped to the UniProt canonical BAF protein isoforms (see Supplementary Table S3). Missense, frameshift, nonsense, and in-frame indels were included, all other mutations were excluded. Additionally, duplicate mutations in patients with multiple samples were excluded. This filtered set of mutations from cBioPortal was used for downstream BAF Cancer vs NDD comparative analyses.

Domains annotations were compiled from PFAM, InterPro and the literature, and manually curated based on the AlphaFold EMBL-EBI structural predictions.

PDB models were obtained from RCSB and PDBDEV (PDB IDs: 6LTJ, 1RYU, 5B79, 7CYU; PDBDEV IDs: PDBDEV_00000056).

Data Availability
Public and private data can be accessed through their respective portals. Private data will require prior authorization. Data can be cleaned and normalized using any standard or well-established procedure for variant analysis or the procedures described in this paper, including referenced papers or procedures. The integrated, curated, and de-duplicated data (to the best of our ability) are available in Supplementary Table 1. No additional data or intermediate results will be available upon request given the high manual burden to verify access to a variety of private portals, repositories and patients. The code used to analysis and generate figures using variant data in Supplementary Table 1 is available under Creative Commons license through Zenodo at (doi: 10.5281/zenodo.8008632). Analyses were executed in Python (v3.7), R (v4.1.1), GraphPad Prism (v92.2), matplotlib(v3.3.1), circos (v0-69-9), and seaborn (v0.11.1).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences   ☐ Behavioural & social sciences   ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	2539 total variants in our dataset, integrated from the following sources: 1) ClinVar, 2) DECIPHER, 3) Literature review, 4) novel, previously-unreported variants, 5) LOVD, Simons Foundation Research Initiative (SFARI) datasets (including 6) SPARK, 7) SSC-ASC) and 8) the Deciphering Developmental Disorders (DDD) study. The collation of all of these data are found in Table S1, which includes source ID, cDNA, protein change, and other information).
Data exclusions	Potential variant duplicates, that could not be confirmed as unique cases, were removed. Non-mSWI/SNF variants within the Coffin-Siris Syndrome Registry dataset were excluded.
Replication	N/A
Randomization	N/A
Blinding	N/A

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒ ☐	Antibodies
☒ ☐	Eukaryotic cell lines
☒ ☐	Palaeontology and archaeology
☒ ☐	Animals and other organisms
☐ ☒	Human research participants
☒ ☐	Clinical data
☒ ☐	Dual use research of concern

Methods

n/a	Involved in the study
☒ ☐	ChIP-seq
☒ ☐	Flow cytometry
☒ ☐	MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	This study analyzed data collected from human participants, however the covariate population characteristics are not included in the data reported as part of this study and were not used by the authors for analysis in this focused analysis. There are currently approximately 500 individuals in the Coffin-Siris syndrome Registry. Most are from the US but individuals from other countries have enrolled as well. Ages range from infancy to middle-age. The sex and age of individuals for whom novel variants are reported in this study are included in Supplementary Table 1.
Recruitment	Recruitment varied by institution. The CSS registry: Eastern Virginia Medical School: Individuals are recruited to the registry through clinicians, social media, and patient foundations. Individuals complete an online consent form followed by a registry survey with phenotypic inquiries. Leiden University Medical Center: No formal recruitment was performed for individuals. Individuals were identified through physician referrals which we have shared with their consent.
Ethics oversight	The Institutional Review Board of Leiden University Medical Center, Leiden, The Netherlands provided approval waivers for using de-identified data and publishing aggregated data (no: G18.098 and G21.129) without obtaining specific informed consent. The Coffin-Siris Syndrome Registry has been approved by the Eastern Virginia Medical School IRB (IRB# 15-03-EX-0058).

Note that full information on the approval of the study protocol must also be provided in the manuscript.