Delineation of phenotypes and genotypes related to cohesin structural protein RAD21

Lianne C. Krab[1,2,*], Iñigo Marcos-Alcalde[3,4], Melissa Assaf[5], Meena Balasubramanian[6], Janne Bayer Andersen[7], Anne-Marie B. Pedersen[8], David R. Fitzpatrick[9], Sanna Gudmundsson[10], Sylvia A. Huisman[1,11], Tugba Kalayci[12], Saskia M. Maas[1,13],Francisco Martinez[14], Shane McKee[15], Leonie A. Menke[1], Jill A. Mokry[16], Paul A. Mulder[17], Oliver D. Murch[18], Michael Parker[19], Juan Pie[20], Feliciano J. Ramos[21], Claudine Rieubland[22], Emanuela Scarano[23], Marwan Shinawi[24], Paulino Gómez-Puertas[3], Zeynep Tümer[7,25,*#], Raoul C. Hennekam[1#]

* corresponding authors

# contributed equally

Author affiliations:

[1] Dept of Pediatrics, Amsterdam UMC, University of Amsterdam, Amsterdam, the Netherlands

[2] Cordaan, Outpatient Clinic for ID Medicine, Amsterdam, and Outpatient Clinic for ID Medicine, Odion, Purmerend, the Netherlands

[3] Molecular Modelling Group, Centro de Biología Molecular Severo Ochoa, CBMSO (CSIC-UAM), Madrid, Spain

[4] School of Experimental Sciences - IIB, Universidad Francisco de Vitoria, UFV, Pozuelo de Alarcón, Spain

[5] Banner Childrens Specialists Neurology Clinic, Glendale, Arizona, USA

[6] Clinical Genetics Service, Sheffield Children's Hospital, and Academic Unit for Child Health, University of Sheffield, Sheffield, UK

[7] Kennedy Center, Department of Clinical Genetics, Copenhagen University Hospital, Rigshospitalet, Glostrup, Denmark

[8] Dept of Pediatrics and Adolescent Medicine, Copenhagen University Hospital, Rigshospitalet, Glostrup, Denmark

[9] MRC Human Genetics, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK

[10] Dept of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden

[11] Prinsenstichting, Purmerend, The Netherlands

[12] Division of Medical Genetics, Dept of Internal Medicine, Istanbul University, Istanbul, Turkey

[13] Department of Clinical Genetics, Amsterdam UMC, University of Amsterdam, Amsterdam, the Netherlands

[14] Unidad de Genética, Hospital Universitario y Politécnico La Fe, Valencia, Spain

[15] Northern Ireland Regional Genetics Service, Belfast City Hospital, Belfast, UK

[16] Dept of Molecular and Human Genetics, Baylor College of Medicine, and Baylor Genetics Laboratories, Houston, Texas, USA

[17] Autism Team Northern-Netherlands, Jonx Dept of Youth Mental Health and Autism, Lentis Psychiatric Institute, Groningen, the Netherlands

[18] Institute of Medical Genetics, University Hospital of Wales, Cardiff, UK

[19] Clinical Genetic Service, Northern General Hospital, Sheffield, UK

[20] Unit of Clinical Genetics Unit, Pediatrics, University Hospital "Lozano Blesa", University of Zaragoza School of Medicine, Zaragoza, Spain

[21] Unit of Clinical Genetics Unitand Functional Genomics, Pediatrics, University Hospital "Lozano Blesa", University of Zaragoza School of Medicine, Zaragoza, Spain

[22] Dept of Pediatrics, Division of Human Genetics, Inselspital, University of Bern, Bern, Switzerland

[23] Rare Disease Unit, Dept of Pediatrics, St Orsola Hospital, Bologna, Italy

[24] Dept of Pediatrics, Division of Genetics and Genomic Medicine, Washington University School of Medicine, St Louis, Missouri, USA

[25] Dept of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark.

Corresponding authors:

Lianne C. Krab, Cordaan, Outpatient Clinic for ID Medicine, Klinkerweg 75, 1033 PK Amsterdam, The Netherlands

Email: lkrab@cordaan.nl Phone: +31-6-24733226

Zeynep Tümer, Kennedy Center, Department of Clinical Genetics, Copenhagen University Hospital, Rigshospitalet, Gl. Landevej 7, 2600 Glostrup

E-mail: Zeynep.tumer@regionh.dk Phone: +45 29204855

Conflict of interests: The Department of Molecular and Human Genetics at Baylor College of Medicine receives revenue from clinical genetic testing conducted at Baylor Genetics Laboratories.

ABSTRACT

Background: *RAD21* encodes a key component of the cohesin complex, and variants in *RAD21* have been associated with Cornelia de Lange Syndrome (CdLS). Limited information on phenotypes attributable to *RAD21* variants and genotype-phenotype relationships is currently published.

Methods: We gathered a series of 49 individuals from 33 families with *RAD21* alterations (40 unique intragenic sequence variants, 7 microdeletions), including 24 hitherto unpublished cases. We evaluated consequences of twelve intragenic variants by protein modelling and molecular dynamic studies.

Results: Full clinical information was available for 29 individuals. Their phenotype is an attenuated CdLS phenotype compared to that caused by variants in *NIPBL* or *SMC1A* for facial morphology, limb anomalies, and especially for cognition and behavior. In the 20 individuals with limited clinical information, additional phenotypes include Mungan syndrome and holoprosencephaly, with or without CdLS characteristics. We describe several additional cases with phenotypes including sclerocornea, in which involvement of the *RAD21* variant is uncertain.

Variants were frequently familial, and genotype-phenotype analyses demonstrated striking interfamilial and intrafamilial variability.

Conclusion: Careful phenotyping is essential in interpreting consequences of *RAD21* variants, and protein modeling and dynamics can be helpful in determining pathogenicity. The current study should be helpful when counseling families with a *RAD21* variation.

Funding: RTC-2017-6494-1, RTI2018-094434-B-I00, European JPIAMR-VRI network “CONNECT”, FIS PI15/00707, DGA-Feder Group B32_17R.

Key words: *RAD21*; *NIPBL*; *SMC1A*; cohesinopathy; Cornelia de Lange Syndrome; Mungan syndrome; genotype-phenotype correlation

INTRODUCTION

RAD21 (ENSG00000164754; OMIM *606462) is a key component of the cohesin complex and it forms a tri-partite ring together with SMC1A and SMC3 (Fig. 1 and Suppl. Fig. S1). The cohesin complex is a major modulator of chromosome structure, is involved in regulating chromosome segregation during mitosis, DNA repair and chromatin condensation, and plays an important role in gene transcription during interphase and cellular homeostasis [1-3]. RAD21 has been implicated in additional processes including mediation of epigenetic silencing and induction of apoptosis [4,5]. Variants in genes encoding various structural or functional components of the cohesin complex, including *RAD21*, *SMC1A, SMC3, BRD4, STAG1/2, NIPBL, HDAC8, WAPL, ANKRD11* and in single individuals *PDS5A* and *ESPL1*, have been implicated in Cornelia de Lange Syndrome (CdLS) [6-9]. *RAD21* spans ~29 Kb and has 14 exons (13 coding, 1 noncoding) that together encode a protein of 631 amino acids [10].

RAD21 variants are found in a minority of CdLS patients. To date, nine missense variants and 5 microdeletions have been reported in CdLS patients [9]. CdLS is characterized by distinct facial features, growth delay, microcephaly, limb reduction defects, intellectual disability (ID) and behavioral problems, especially self-injurious behavior (SIB) and autism spectrum disorder (ASD) [9]. *RAD21* variants have also been associated with sclerocornea [11] and Mungan syndrome (OMIM #611376) [12,13], each in a single family in which no remarks on CdLS features were made in the report. Loss of function-variants in cohesin genes including *RAD21* were found in individuals with holoprosencephaly of whom some demonstrated CdLS features as well [14].

RAD21 is positioned on chromosome 8q24.11, between *TRPS1* (Tricho-Rhino-Phalangeal syndrome type 1; OMIM *604386) and *EXT1* (Multiple Exostoses type 1; OMIM *608177). Several microdeletions involving *RAD21* encompass genes next to *RAD21* (contiguous gene syndrome), complicating attribution to *RAD21* of the phenotype [15-17]. TRPS type 2 or Langer-Giedion syndrome (OMIM #150230) involves *TPRS1, RAD21* and *EXT1*, and the facial phenotype is mainly determined by loss of *TRPS1*, whereas the bony abnormalities arise from loss of *EXT1* [18].

Based on the small case series of CdLS patients with *RAD21* variants reported so far, face and limb manifestations of CdLS seem to be less pronounced compared to individuals with variants in the other cohesin complex genes, and the impact on cognitive functioning seems attenuated, without clear genotype-phenotype correlation [9,19]. Here, we report on a case series of 49 patients from 33 families with *RAD21* alterations, including all previously published cases with sequence variants, most of which with updated clinical data. We included 24 hitherto unpublished cases. We present genotype data, evaluate pathogenicity of intragenic variants by a combination of phenotype, protein modelling, and molecular dynamic studies, and provide information on clinical phenotype, including cognitive and behavioral functioning, interfamilial and intrafamilial variability, and genotype-phenotype associations. We compare the *RAD21* phenotype to that of patients with *NIPBL* and *SMC1A* variants.

RESULTS

We identified 219 cases with *RAD21* variants, of which 49 patients from 33 families were included in this study (Table S1). We describe in Table S6 those excluded cases that still may be of interest such as published cases with involvement of other morbid genes [15-18,23], variants reported as variant of unknown significance (VUS) that remained with unknown significance subsequent to re-evaluation, and cases for whom the relationship between phenotype and *RAD21* variant could not be confirmed [11,14].

The 49 patients can be divided into two groups: cohort A includes 29 patients (22 families) with sufficient clinical data; and cohort B includes 20 patients (11 families) with incomplete data. Of the 49 cases, 24 are new. Twenty-five were previously published [6,7,12,14,17,19,24-29], and for 19 of these clinical data could be updated (Table 1). Patients originated from Australia, Belgium, Canada, Denmark, Germany, Italy, Netherlands, Spain, Sweden, Switzerland, Turkey, United Kingdom and United States.

Genotype

The 33 families harbor 29 unique and 2 recurrent variants [p.(Cys585Arg) and p.(Arg586*), each found in 2 families (Table 1, Fig. 1)]. A relatively large proportion of the cases are familial (nine out of 21 index cases for whom inheritance could be established). Seven variants were copy number variations (CNVs): deletions of which six include genes besides *RAD21*. All CNVs where segregation could be investigated are *de novo*. Of the 24 unique sequence variants 13 are predicted be truncating (2 nonsense, 2 splice site and 9 frameshift variants), which are scattered throughout the gene. Three of the variants are in-frame deletions, two of which affect a single amino acid, while the 665 bp deletion includes the whole exon 13. Six out of eight unique missense variants are clustered at the functional domains of the protein. Some variants in cohort B may be recurrent but sufficient data are lacking to confirm this (Table S6).

Evaluation of pathogenicity of *RAD21* variants using molecular dynamic analyses

For twelve intragenic variants (ten missense variants and two 3bp in frame deletions, from individuals in cohort A, B and Table S6) it was possible to carry out structural analysis, as their substituted residues are located in one of the domains for which 3D arrangement can be modeled (RAD21-SMC3 domain, RAD21-STAG domain and RAD21-SMC1A domain, Fig. 2; Figs. S2-3). Interactions between RAD21 and its binding partners are shown in Fig. S1.

RAD21-SMC3 domain (residues 18-87) harboring Arg65Gln, and RAD21-STAG domain (residues 321-392) harboring Ser345Pro, Pro355Leu and Pro376Arg: Substitution of Arg65 with Gln (Arg65Gln) is a semi-conservative variation that did not promote detectable structural or dynamic changes in the complex but could be partially disrupting a putative interaction with DNA. Ser345Pro variant impairs RAD21 and STAG1/2 interactions due to promotion of a *de novo* curved small alpha helix segment that binds to the pre-existing alpha helix, which separates from the surface of STAG2. No structural or dynamic effects of Pro355Leu or Pro367Arg on RAD21 itself could be observed. Nevertheless, Pro376Arg does promote formation of a new salt bridge between RAD21 and STAG2, which is predicted to cause over-stabilization of the interaction between the two proteins.

RAD21-SMC1A domain (residues 543-628) harboring Gly575Ala, Cys585Arg, Arg586Gln, Gln592del, Phe600del, Leu603Pro, Ser618Gly, and Ala622Thr: Four of the eight variants in this domain (Cys585Arg; Arg586Gln; Gln592del; Leu603Pro) are predicted to cause a structural effect. Arg586Gln destabilizes the RAD21-SMC1A domain by loss of a salt bridge between Arg586 and Glu577, and the altered position of Glu577 adds an additional negative charge to the RAD21 surface of RAD21-SMC1A. Cys585Arg has a similar effect, interacting with Glu583 and causing Arg586 to lose its contact with Glu577. The MD simulation shows that both Gln592del and Leu603Pro, but not Phe600del, affect positioning of SMC1A-Asn35 at the ATPase site 1 by changing the position of Lys605.

Phenotype

Physical features: Individual CdLS scores and major and minor anomalies in cohort A are provided in Table S2-3. Clinical features of cohort A are compared to those of *NIPBL* and *SMC1A* cohorts in Table 2 and illustrated in Fig. 3 and S4. Clinical information for cohort B is available in supplemental materials S5 and will not be discussed further in the text, as clinical data are limited. We mention data in the text only if not represented in the tables.

All patients in cohort A (age range 0-61 years, median nine years, mean 18 years; 15 males) had CdLS scores of at least five, sufficient to warrant molecular genetic testing for CdLS. In about 60 percent of index cases (13/21 index cases in which this was specified) CdLS was suspected prior to testing. There was no gender difference in CdLS scores. No *RAD21* variant would have been missed by using the CdLS consensus criteria for molecular studies [9]. Clinical scores of patients with CdLS suspected prior to testing (median 11.5; range 8-13) were higher than those not suspected to have CdLS (median 9.5; range 5-13).

Cognition, development and behavior*:* Cognitive functioning, developmental milestones and behavioral functioning in the *RAD21* group are attenuated compared to the *NIPBL* and *SMC1A* groups (Tables 3 and S4). The majority of *RAD21* patients (16/29, 55%) have normal or mildly impaired cognitive functioning (*SMC1A* group 32%; *NIPBL* group 7%) [20,21]. In all three groups there is a trend towards more language-based problems than motor-based problems in development. Still, all *RAD21* patients aged 3 years and above were able to use some words. There was no correlation between the severity of cognitive impairment in *RAD21* patients and presence of microcephaly (prenatal, postnatal, or both; data not shown).

14/25 *RAD21* patients (56%) with sufficiently available information on behavior had problems, mainly features of anxiety, ADHD, ASD, and obsessive-compulsive behavior. ASD related problems, aggression and SIB were less prevalent compared to the *SMCIA* and *NIPBL* groups.

Genotype-phenotype comparisons in cohort A

***Microdeletions versus intragenic variant*s:** There was a trend towards higher CdLS scores and more frequently impaired growth parameters in patients with microdeletions compared to those with intragenic variants, but no differences were apparent in frequency of major malformations or cognitive and behavioral problems. We refrained from statistical analyses as small numbers would make results too unreliable and less useful. Exostoses, related to *EXT1* haploinsufficiency, likely caused the upper limb anomalies.

Truncating versus non-truncating sequence variants: There was no difference in CdLS scores or growth parameters between individuals with truncating and those with non-truncating sequence variants (median 10; range 9-13 and median 9.5; range 5-12, respectively).

Malformations and genotype: For 12/15 patients with intragenic variants and major malformations or health problems, the variant was located in a protein-binding domain (F2, F3a, F8, F9, F11a, F11b, F12, F14a, F14b, F16a, F17, F18). As numbers are small it remains uncertain whether this is truly an association. The types of major malformations did not differ.

***Intrafamilial variation*:** The intrafamilial variation can be considerable (Tables S1, S3-4; Fig. 3), especially in cognition and behavior. Through obvious ascertainment bias cognition is more frequently impaired in index cases. Several families include patients with ID and patients with apparently normal cognitive functioning. The intrafamilial variation cannot be explained by mosaicism in most families.

DISCUSSION

We report on *RAD21* variants in 49 individuals, some with sufficient clinical data (cohort A), others with limited clinical data (cohort B). *RAD21* variants are frequently familial, often unique, and without obvious hotspots for variants or microdeletions breakpoints, although missense variants tend to cluster around protein binding domains.

RAD21 missense variants and their predicted effect on protein function

Combining our data from structural and functional analysis with data from phenotype, database information and literature, nine out of twelve *RAD21* missense variants that we modelled were demonstrated to be likely pathogenic. The structural and dynamic analyses indicate the Arg65Gln variant within the RAD21-SMC3 domain does not make it likely to be pathogenic. However, Arg65 is located in the close proximity of Tyr67, and altering the kinase/phosphatase recognition motif Arg-X-Tyr around Tyr67 may affect the phosphorylation-based regulation of RAD21 [30-33]. In addition, a contact between the PDS5 protein and the RAD21-SMC3/SMC3-head complex is involved in the topological entrapment of DNA by cohesin [34]. As Arg65 is located towards the solvent, Arg65Glu may impact the RAD21-PDS5 recognition and, thus, disturb their interaction.

The interaction between RAD21 and STAG1/2 is crucial for the proper functioning of the cohesin complex [34], and both impairing (Ser345Pro) or over-stabilizing (Pro367Arg) variants within the RAD21-STAG domain are predicted to cause dysfunction of the complex, presumably through affecting the continuous cycle of formation and disengagement of the cohesin ring [35].

The structural model of the RAD21-SMC1A domain rationalizes the key function of RAD21 in the ATPase reaction at the SMC1A/SMC3 head, which is pivotal to the opening of the cohesin ring and, thus, the cyclic process [35]. The Cys585Arg and Arg586Gln variants destabilize the RAD21-SMC1A domain; and Gln592del and Leu603Pro (but not Phe600del) disturb the cyclic process through dislocation of Lys605. However, as the Phe600del variant does lead to a classical CdLS phenotype without variants in additional known CdLS genes, it does seem likely pathogenic. Unfortunately, the

crystal structure of RAD21 is not available for other domains or interacting partners such as WAPL and PDS5, but earlier molecular studies provide additional information for other missense variants.

The importance of the regulation of interaction between RAD21-SMC1A and SMC1A/SMC3 head is demonstrated by the several residues involved in phosphorylation and ubiquitination in the RAD21-SMC1A domain [31,32,36]. Ala622 is positioned next to Thr623, a substrate for protein phosphorylation by PLK1 [32,37]. A pathogenic effect of variant Ala622Thr is supported by studies showing decreased bowel transit and loss of enteric neurons in zebrafish with Ala622Thr variants and patients with Mungan syndrome and CIPO (chronic intestinal pseudo-obstruction) [12].

Two additional variants outside the domains that can be modeled, are likely to be pathogenic. As residue Thr461 is flanked by Ser residues (Ser459 and Ser466), both implicated in phosphorylation-regulated dissociation of cohesin from chromosome arms [32,38], it may modify the kinase/phosphatase recognition motif, thus affecting the protein behavior. Similarly, Phe6 is found close to Ser9, a phosphorylation site described in the human proteome [34,39]. The Phe6Val variant would modify the kinase/phosphatase recognition motif.

Clinical phenotype

Physical phenotype: *RAD21* variants can lead to a CdLS phenotype (*RAD21*-CdLS). The (limited) available information of individuals from cohort B suggests that *RAD21* variants can also lead to CIPO and (like one case in cohort A) holoprosencephaly and possibly schizophrenia, although in the latter the association may be a spurious coincidence. In Table S6 we describe several additional cases with phenotypes including sclerocornea and schizophrenia, in which pathogenicity of the *RAD21* variant is debatable. Due to incomplete information it remains uncertain whether these individuals are also showing CdLS characteristics. Indeed, when we succeeded in obtaining further clinical information, several individuals turned out to show CdLS characteristics not mentioned in the publication (for instance in the family with CIPO). Additionally, one may speculate that phenotypes are also attributable (possibly in addition to the *RAD21* variant) to variants in other genes.

Comparison to phenotypes of NIPBL and SMC1A variants: In patients with sufficient clinical data available (cohort A) most features associated with CdLS are present. However, the prevalence of features is lower compared to those in the *SMC1A* and *NIPBL* cohort, and the degree of severity is typically less. Severe visual impairment and diaphragmatic hernias are rare in *RAD21* patients, and feeding difficulties are uncommon. *RAD21* patients less frequently have increased body hair (hirsutism, bushy eyebrows, low scalp hair lines), major limb malformations are not reported, and hands and feet are generally of normal size. Still, minor anomalies of hands and feet are common, such as fetal pads, abnormal flexion crease patterns, and camptodactyly. Patients with *RAD21* variants have generally less impaired growth at birth, and short stature and microcephaly develop postnatally. Prenatal microcephaly has been demonstrated to be a predictor of more severe cognitive impairment in CdLS in the pre-molecular era, [40] but this does not hold for *RAD21* patients. Frequency and severity of congenital heart defects is similar to those in the *NIPBL* and *SMC1A* cohorts. Gastro-esophageal reflux is similar in frequency but in *RAD21* it is typically mild and restricted to early childhood. No *RAD21* patients exhibit a Rett-like phenotype as can occur in *SMC1A* patients [20]. The CdLS score remains a reliable tool, and the present study does not call for adjustment of the diagnostic advice from the CdLS guidelines [9].

Unusual anomalies in the *RAD21* cases are vertebral anomalies (clefts and hemivertebrae). There is a single individual with a *NIPBL* variant and Klippel-Feil anomaly (personal observation RCH), and upper cervical spine malformations have been reported in other patients with *NIPBL* variants as well [41]. Malformations of structures derived from the embryonic foregut are relatively frequent in *RAD21* patients and have only rarely been described in CdLS [42-44]. Holoprosencephaly spectrum anomalies have been linked to several cohesin genes [14], including *RAD21*, although in one individual this remains uncertain (Table S6). The prevalence of holoprosencephaly spectrum in *RAD21*-CdLS must remain uncertain as brain MRIs are typically not indicated in individuals with CdLS due to the burden of the procedure and lack of consequences of findings for care [9].

Development, cognition and behavior: Most data on cognition and behavior in the present cohort are based on subjective information provided by physicians and not on formal testing. Therefore, reliability remains uncertain. Still, all data point to a lower prevalence and decreased severity of ID in *RAD21* patients compared to *NIPBL* and *SMC1A* groups: developmental milestones are more frequently attained, cognitive level is estimated higher, aggression less and autism less frequent. SIB, a hallmark of CdLS in general [9], is infrequent in *RAD21* individuals.

Also, if an IQ is normal, subtle difficulties in neuropsychological domains known to be affected in CdLS [9] may influence cognitive performance. Periodic formal screening for neuropsychological and behavioral problems is still warranted in all individuals with *RAD21* variants, to allow for early recognition of problems and access to relevant support systems. In addition, formal (in-person) assessments can prevent misdiagnoses, such as autism, by putting behavioral characteristics into perspective of the developmental level of patients [21].

Natural history*:* The natural history data from the present study indicate that pregnancies and birth tend to progress normally, prenatal growth retardation being present in a small minority. About half of the patients have congenital anomalies (cleft palate; cardiac anomalies). Major limb defects have not been found; diaphragmatic hernia, anal atresia or choanal atresia occur occasionally. Patients have typically mild facial dysmorphisms, no small hands or feet, and increased body hair is less apparent compared to *SMC1A* and *NIPBL* patients. The clinical diagnosis of CdLS may therefore be difficult.

Neonatal feeding is usually not problematic. Reflux is common but not severe. Typical development is somewhat slow, mainly in speech development, and physical therapy or speech therapy may be indicated. As they grow up, children only occasionally develop new medical problems. Half of the children show a progressive but still mild growth delay in head circumference and height. Vision is mostly normal; hearing loss is found in a third of individuals and may require

hearing devices. Most of the patients are able to attend regular education or education for children with mild cognitive disabilities. Most have some behavioral problems (mainly anxiety, ADHD or ASD) of limited severity, and aggression and SIB are uncommon. Not uncommonly, *RAD21* patients are able to start a family, and some are only diagnosed when more severely affected offspring is recognized.

Genotype-phenotype associations: Patients with haploinsufficiency for *RAD21* due to microdeletions or truncating variants do not differ markedly from those with missense variants, and nonsense mediated decay is not apparent. It remains uncertain whether duplication of the whole gene can lead to a CdLS phenotype, as demonstrated for duplications in *STAG2* and *SMC1A* [45,46], as all duplications we retrieved, were either including several other genes or pathogenicity could not be confirmed. No fully intragenic duplication is known to us. Small duplications have also been detected in apparently healthy controls (unpublished observations J. Howe).

In general, protein studies combined with a detailed phenotype allow often, but not always, to probe for RAD21 dysfunction in patients with variants of doubtful meaning. The phenotype in individuals with *RAD21* variants is not only determined by the variant itself but potentially also by other factors: 1) variable expression of cohesin subunits and/or binding partners in different tissues; 2) variable formation of isoforms in different tissues; 3) modifying genes, especially of the cohesin complex [6]; 4) Epigenetic factors such as methylation and gene silencing [47]. Exact phenotypic consequences, if any, of each of the above are unknown.

In counseling of families with *RAD21* variants, the relatively high frequency of familial occurrence and marked intrafamilial and interfamilial variability should be considered. Parental testing is warranted, even if signs or symptoms are apparently absent in parents, and standard testing of parents may further broaden the phenotype of *RAD21* variants. We suggest a cautious use of data on variants in molecular databases, as due to the extremely variable and sometimes very mild phenotype wrong conclusions may be drawn in classifying the variants. In case of a CdLS phenotype

and detection of a VUS in *RAD21*, we recommend testing for variants in other CdLS associated genes and eventually carry out 'open' exome/genome sequencing.

Limitations: Although we used a broad search strategy and the present *RAD21* cohort is the largest reported thus far, numbers are still small, and these preclude further statistical analyses. We did not consider variants from ClinVar or Decipher that were reportedly (likely) benign, but we expect that these may contain some pathogenic variants discarded based on an overlooked (mild) phenotype. Furthermore, many variants are reported with insufficient clinical data preventing such patients to be included in the present series. Especially morphological data are often missing, and we stress the importance of the use of the CdLS consensus data in evaluating individuals with variants in cohesin genes [9]. Next generation sequencing based technologies such as gene panels or 'open' exome/genome sequencing remains to be introduced in many countries, and we expect identification of many additional patients with pathogenic *RAD21* variants as clinical recognition may be difficult. Lastly, we may have an acquisition bias due to involvement of specialists in CdLS, causing an overrepresentation of individuals with a CdLS phenotype.

Future: The present results demonstrate that more information on larger groups of individuals with *RAD21* variants is needed to determine the complete phenotypic spectrum. CdLS characteristics such as sleep disturbances and autonomic dysfunctions in individuals with *RAD21* variants are still largely unknown. A specific issue that needs attention is the risk to develop cancer (not reported to date in RAD21 patients) [17,19]. We call also for more detailed study of cognitive, behavioral and psychiatric phenotypes, as these are of utmost importance in clinical care. Molecular and cellular mechanisms underlying cognitive problems are unclear, although cohesin-mediated 3D-organization of the genome suggests a role for neuronal plasticity [48]. Studying RAD21 and other cohesin components in this process could contribute to the search for targeted influencing of cognition and especially behavior in CdLS. Effects of *RAD21* variants on cellular functioning and relationships between

genotype and phenotype may be elucidated further by studying genome-wide methylation patterns (episignature), which have been shown to be altered in CdLS [49]. This may explain presently unexpected discrepancies between genotype and phenotype, and even allow for establishing pathogenicity in individuals with uncertain molecular findings.

METHODS

Patients

Patients were gathered using a combination of literature and database search and network inquiries (see Supporting Information). A dedicated questionnaire was used to gather clinical, molecular, cognitive and behavioral data. If allowed by the family clinical pictures were gathered for scoring of facial characteristics by the senior author (RCH). If no clinical pictures were available to us (n=3 in cohort A) the clinician-reported description of facial characteristics was accepted. The CdLS clinical score (reflecting the similarity of clinical features to those in classical CdLS) was computed using cardinal features (2 points each) and suggestive features (1 point each) according to Kline and co-authors [9].

Information on cognitive functioning and behavioral problems was derived from physician reported data, if possible substantiated with results of formal testing. For the CdLS clinical score, minor criterion "ID or global DD" was scored positive if ID or global DD (global developmental delay; a combination of delay in at least 2 developmental domains) was present, at any age. Elsewhere in the manuscript, cognitive functioning has been classified into categories based on DSM-5.

To compare the *RAD21* phenotype to CdLS patients with variants in other genes, clinical data were obtained from existing *NIPBL* and *SMC1A* cohorts [20,21], to which we added further information if needed. For comparison of features of ASD and aggression in *NIPBL* patients, we derived information from the UK cohort [22]. For the item 'autistic like behavior', we compared with scores from the Social Communication Questionnaire (number above cut-off for ASD); for the item 'aggression' with presence of verbal aggression, physical aggression or property destruction on the Challenging Behavior Questionnaire; and for 'obsessive-compulsive behavior' with the number of patients with one or more items of the compulsive behavior subscale of the Repetitive Behavior Questionnaire above clinical cutoff.

Based on availability of clinical data, we composed two cohorts: cohort A with sufficient clinical data available on all cardinal CdLS features, and cohort B with incomplete clinical data. We provide an overview of excluded cases in Supporting Information Table S6.

Molecular studies

Among the 29 patients (22 index) of cohort A, a clinical diagnosis of CdLS was suspected in 13 index cases, which allowed detection of *RAD21* variants using array comparative genomic hybridization (CGH, (n=1), Sanger sequencing (n=6), 'whole' exome sequencing (WES, n=2), or targeted exome sequencing searching for variants in genes that can cause intellectual disability (ID-WES, n=4). Confirmation by Sanger sequencing of an exome result was only performed if the coverage of the exome was thought to be of insufficient quality. The other nine index cases were detected through Sanger sequencing of a series of candidate genes after excluding a clinical diagnosis (KBG syndrome, n=1), or after WES (n=2), ID-WES (n=3), or array CGH or SNP array (n=3). All molecular studies were performed for diagnostic reasons, following the various national regulations, and for none of the patients studies were performed because of the current research. For describing the variants coding DNA reference sequence NM_006265.2(*RAD21*_v001) is used.

Structure modeling of RAD21 variants

A set of three wild-type and twelve variant protein models was generated through standard homology modeling procedures using the SWISS-MODEL server (http://swissmodel.expasy.org; see Supporting Information). These were used to study the structural effects of the missense variants located in the protein domains in contact with SMC3 (RAD21 N-terminus; RAD21-SMC3), STAG1/2 (RAD21-STAG) or SMC1A (RAD21 C-terminus, RAD21-SMC1A).

Molecular dynamics simulations

To analyze the putative effect of variants on the RAD21 structure, the behavior of the twelve variant proteins were compared to that of wild type models by free molecular dynamics (MD) simulation for 60-100 nanoseconds (ns) (see Supporting Information). Movements during the trajectories were continuously measured by root-mean square deviation (RMSD) of atomic positions. Large variations of RMSD values indicate notable distortions of protein structure due to the abnormal amino acid variant. RAD21 domains were modeled in complex with the accompanying proteins, to facilitate functional evaluation of variants along the MD trajectories.

Ethics

All clinical investigation has been conducted according to Declaration of Helsinki principles. Written informed consent was received from participants prior to inclusion in the study. Patients or their legal representatives have provided written consent for using images. The Medical Ethics Committee of the Amsterdam UMC approved the study (NL39553.018.12).

AUTHOR CONTRIBUTIONS

LCK, ZT and RCH designed the study. RH, LCK and ABP acquired data. LCK, ZT and RCH analyzed the data. IM-A and PG-P performed the protein studies. LCK, ZT, PG-P and RCH wrote the manuscript. MA, MB, JBA, AB, KC, DF, SG, SMcK, SMM, LAM, FM, JAM, ODM, MP, JP, FJR, CR, ES, MS and RCH provided (updated) clinical data for one or more cases. SH, PAM and RCH provided additional data for the *NIPBL* and *SMC1A* patient cohorts. SH, PAM, SM and LAM contributed to fruitful discussions. All authors revised the manuscript before submission.

ACKNOWLEDGEMENTS

We thank Drs. Wait Kit Chu, Cheryl Greenberg, Jennifer Howe, Frank Kaiser, Chitra Prassad, and Mrs M. Towne for their efforts to retrieve additional data on cases. We thank of Hansol Lee, Byung-Ha Oh and Stephan Gruber for allowing use of their *pyrococcus yayanosii* SMC ring structure as a template. This work was supported by grants from the Spanish Ministry of Science, Innovation and Universities/State Research Agency RTC-2017-6494-1 and RTI2018-094434-B-I00 (MCIU/AEI/FEDER, UE) as well as funds from the European JPIAMR-VRI network “CONNECT” to PG-P. The computational support of the "Centro de Computación Científica CCC-UAM" is gratefully recognized. This work is generated within the European Reference Network ITHACA.

REFERENCES

1. Watrin E, Kaiser FJ, Wendt KS. Gene regulation and chromatin organization: relevance of cohesin mutations to human disease. *Curr Opin Genet Dev.* 2016;37:59-66.
2. Kamada K, Barilla D. Combing Chromosomal DNA Mediated by the SMC Complex: Structure and Mechanisms. *Bioessays.* 2018;40(2).
3. Mullenders J, et al. Cohesin loss alters adult hematopoietic stem cell homeostasis, leading to myeloproliferative neoplasms. *J Exp Med.* 2015;212(11):1833-1850.
4. Fisher JB, et al. The cohesin subunit Rad21 is a negative regulator of hematopoietic self-renewal through epigenetic repression of Hoxa7 and Hoxa9. *Leukemia.* 2017;31(3):712-719.
5. Pati D, Zhang N, Plon SE. Linking sister chromatid cohesion and apoptosis: role of Rad21. *Mol Cell Biol.* 2002;22(23):8267-8277.
6. Yuan B, et al. Clinical exome sequencing reveals locus heterogeneity and phenotypic variability of cohesinopathies. *Genet Med.* 2019;21(3):663-675.
7. Ansari M, et al. Genetic heterogeneity in Cornelia de Lange syndrome (CdLS) and CdLS-like phenotypes with observed and predicted levels of mosaicism. *J Med Genet.* 2014;51(10):659-668.
8. Woods SA, et al. Exome sequencing identifies a novel EP300 frame shift mutation in a patient with features that overlap Cornelia de Lange syndrome. *Am J Med Genet A.* 2014;164A(1):251-258.
9. Kline AD, et al. Diagnosis and management of Cornelia de Lange syndrome: first international consensus statement. *Nat Rev Genet.* 2018;19(10):649-666.
10. McKay MJ, et al. Sequence conservation of the rad21 Schizosaccharomyces pombe DNA double-strand break repair gene in human and mouse. *Genomics.* 1996;36(2):305-315.
11. Zhang BN, et al. A Cohesin Subunit Variant Identified from a Peripheral Sclerocornea Pedigree. *Dis Markers.* 2019;2019:8781524.

12. Bonora E, et al. Mutations in RAD21 disrupt regulation of APOB in patients with chronic intestinal pseudo-obstruction. *Gastroenterology.* 2015;148(4):771-782 e711.

13. Mungan Z, et al. Familial visceral myopathy with pseudo-obstruction, megaduodenum, Barrett's esophagus, and cardiac abnormalities. *Am J Gastroenterol.* 2003;98(11):2556-2560.

14. Kruszka P, et al. Cohesin complex-associated holoprosencephaly. *Brain.* 2019;142(9):2631-2643.

15. Wuyts W, et al. Multiple exostoses, mental retardation, hypertrichosis, and brain abnormalities in a boy with a de novo 8q24 submicroscopic interstitial deletion. *Am J Med Genet.* 2002;113(4):326-332.

16. Pereza N, et al. Third case of 8q23.3-q24.13 deletion in a patient with Langer-Giedion syndrome phenotype without TRPS1 gene deletion. *Am J Med Genet A.* 2012;158A(3):659-663.

17. Deardorff MA, et al. RAD21 mutations cause a human cohesinopathy. *Am J Hum Genet.* 2012;90(6):1014-1027.

18. Maas SM, et al. Phenotype and genotype in 103 patients with tricho-rhino-phalangeal syndrome. *Eur J Med Genet.* 2015;58(5):279-292.

19. Minor A, et al. Two novel RAD21 mutations in patients with mild Cornelia de Lange syndrome-like presentation and report of the first familial case. *Gene.* 2014;537(2):279-284.

20. Huisman S, et al. Phenotypes and genotypes in individuals with SMC1A variants. *Am J Med Genet A.* 2017;173(8):2108-2125.

21. Mulder PA, et al. Development, behaviour and autism in individuals with SMC1A variants. *J Child Psychol Psychiatry.* 2019;60(3):305-313.

22. Moss J, et al. Genotype-phenotype correlations in Cornelia de Lange syndrome: Behavioral characteristics and changes with age. *Am J Med Genet A.* 2017;173(6):1566-1574.

23. Yuen RK, et al. Whole-genome sequencing of quartet families with autism spectrum disorder. *Nat Med.* 2015;21(2):185-191.

24. Boyle MI, Jespersgaard C, Nazaryan L, Bisgaard AM, Tumer Z. A novel RAD21 variant associated with intrafamilial phenotypic variation in Cornelia de Lange syndrome - review of the literature. *Clin Genet.* 2017;91(4):647-649.

25. Dorval S, et al. A novel RAD21 mutation in a boy with mild Cornelia de Lange presentation: Further delineation of the phenotype. *Eur J Med Genet.* 2019;62:103526.

26. Gudmundsson S, et al. A novel RAD21 p.(Gln592del) variant expands the clinical description of Cornelia de Lange syndrome type 4 - Review of the literature. *Eur J Med Genet.* 2018.

27. McBrien J, et al. Further case of microdeletion of 8q24 with phenotype overlapping Langer-Giedion without TRPS1 deletion. *Am J Med Genet A.* 2008;146A(12):1587-1592.

28. Martinez F, et al. High diagnostic yield of syndromic intellectual disability by targeted next-generation sequencing. *J Med Genet.* 2017;54(2):87-92.

29. Lee H, et al. Clinical exome sequencing for genetic identification of rare Mendelian disorders. *JAMA.* 2014;312(18):1880-1887.

30. Li H, et al. Identification of tyrosine-phosphorylated proteins associated with metastasis and functional analysis of FER in human hepatocellular carcinoma cells. *BMC Cancer.* 2009;9:366.

31. Hoque MT, Ishikawa F. Human chromatid cohesin component hRad21 is phosphorylated in M phase and associated with metaphase centromeres. *J Biol Chem.* 2001;276(7):5059-5067.

32. Hornbeck PV, et al. PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res.* 2015;43(Database issue):D512-520.

33. Amanchy R, et al. Identification of Novel Phosphorylation Motifs Through an Integrative Computational and Experimental Analysis of the Human Phosphoproteome. *J Proteomics Bioinform.* 2011;4(2):22-35.

34. Guacci V, Chatterjee F, Robison B, Koshland DE. Communication between distinct subunit interfaces of the cohesin complex promotes its topological entrapment of DNA. *Elife.* 2019;8.

35. Marcos-Alcalde I, et al. Two-step ATP-driven opening of cohesin head. *Sci Rep.* 2017;7(1):3266.

36. Hegemann B, et al. Systematic phosphorylation analysis of human mitotic protein complexes. *Sci Signal.* 2011;4(198):rs12.

37. Tsai CF, et al. Large-scale determination of absolute phosphorylation stoichiometries in human cells by motif-targeting quantitative proteomics. *Nat Commun.* 2015;6:6622.

38. Hauf S, et al. Dissociation of cohesin from chromosome arms and loss of arm cohesion during early mitosis depends on phosphorylation of SA2. *PLoS Biol.* 2005;3(3):e69.

39. Gauci S, et al. Lys-N and trypsin cover complementary parts of the phosphoproteome in a refined SCX-based approach. *Anal Chem.* 2009;81(11):4493-4501.

40. Hawley PP, Jackson LG, Kurnit DM. Sixty-four patients with Brachmann-de Lange syndrome: a survey. *Am J Med Genet.* 1985;20(3):453-459.

41. Bettini LR, et al. Cervical spine malformation in cornelia de lange syndrome: a report of three patients. *Am J Med Genet A.* 2014;164A(6):1520-1524.

42. Mende RH, Drake DP, Olomi RM, Hamel BC. Cornelia de Lange Syndrome: A Newborn with Imperforate Anus and a NIPBL Mutation. *Case Rep Genet.* 2012;2012:247683.

43. Hamilton J, Clement WA, Kubba H. Otolaryngological presentations of Cornelia de Lange syndrome. *Int J Pediatr Otorhinolaryngol.* 2014;78(9):1548-1550.

44. Kang MJ, Ahn SM, Hwang IT. A Novel Frameshift Mutation (c.5387_5388insTT) in NIPBL in Cornelia de Lange Syndrome with Severe Phenotype. *Ann Clin Lab Sci.* 2018;48(1):106-109.

45. Mullegama SV, et al. Mutations in STAG2 cause an X-linked cohesinopathy associated with undergrowth, developmental delay, and dysmorphia: Expanding the phenotype in males. *Mol Genet Genomic Med.* 2019;7(2):e00501.

46. Baquero-Montoya C, et al. Could a patient with SMC1A duplication be classified as a human cohesinopathy? *Clin Genet.* 2014;85(5):446-451.

47. Aref-Eshghi E, et al. Diagnostic Utility of Genome-wide DNA Methylation Testing in Genetically Unsolved Individuals with Suspected Hereditary Conditions. *Am J Hum Genet.* 2019;104(4):685-700.

48. Fujita Y, Yamashita T. Spatial organization of genome architecture in neuronal development and disease. *Neurochem Int.* 2018;119:49-56.

49. Aref-Eshghi E, et al. Evaluation of DNA methylation EpiSigns for diagnosis and phenotype correlations in 42 Mendelian neurodevelopmental disorders *submitted.*

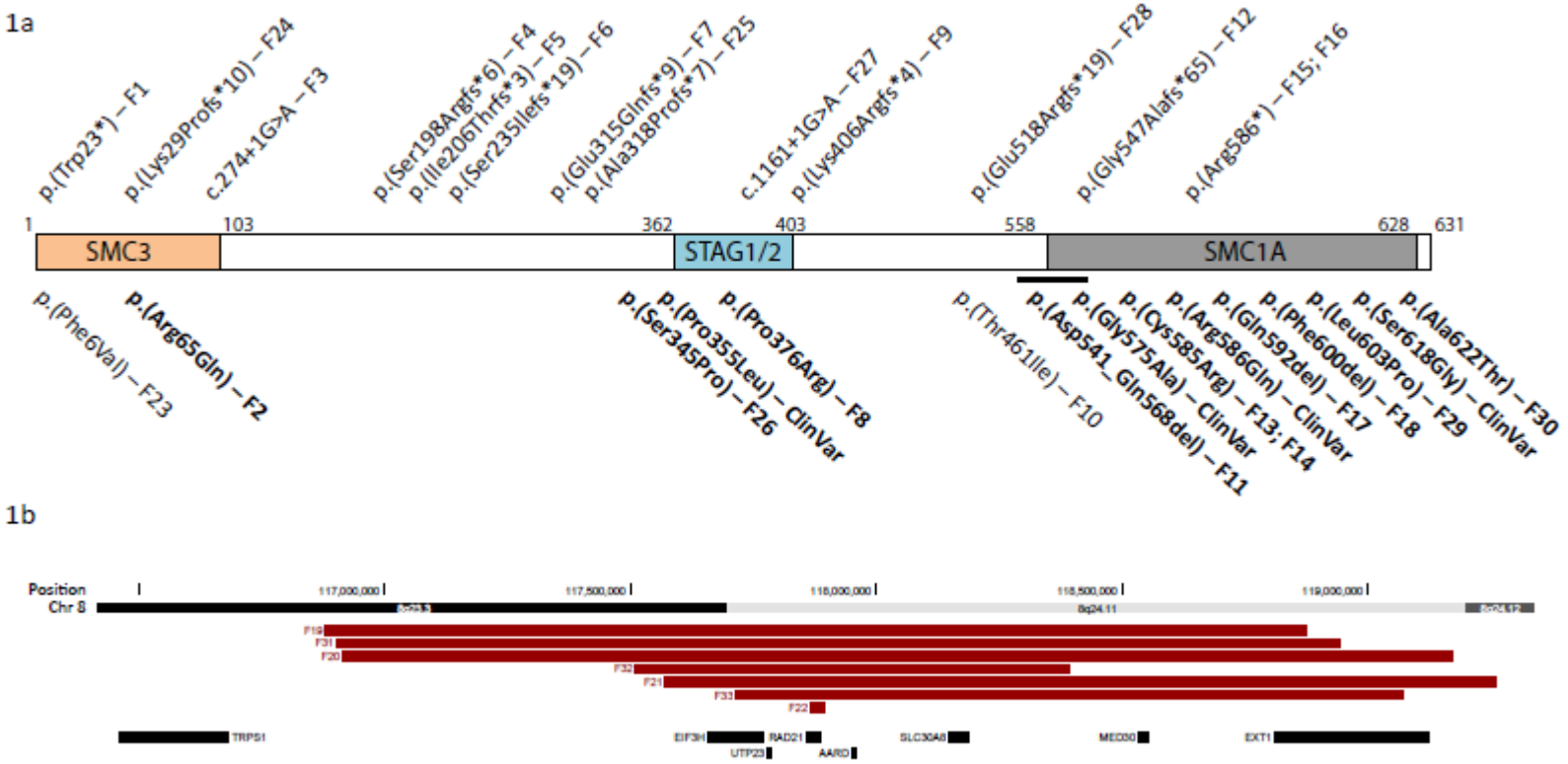

Figure 1 A-B. Presently reported *RAD21* variants. 1a. RAD21 (horizontal bar) has three binding domains: SMC3 (p.1–103), STAG1/2 (p.362–403) and SMC1A (p.558–628). Sizes of the binding domains are not shown to scale. Truncating *RAD21* variants are shown above, and missense mutations and in-frame deletions are shown below the protein representation. Variants for which protein modelling is available, are marked in bold. F: family number. The horizontal black line represents the inframe deletion p.(Asp541_Gln568del). ClinVar, variants which are reported in the ClinVar database and could be investigated for pathogeneity with protein modelling (see supplementary Table S6). **1b.** Genomic region showing the microdeletions including *RAD21*.

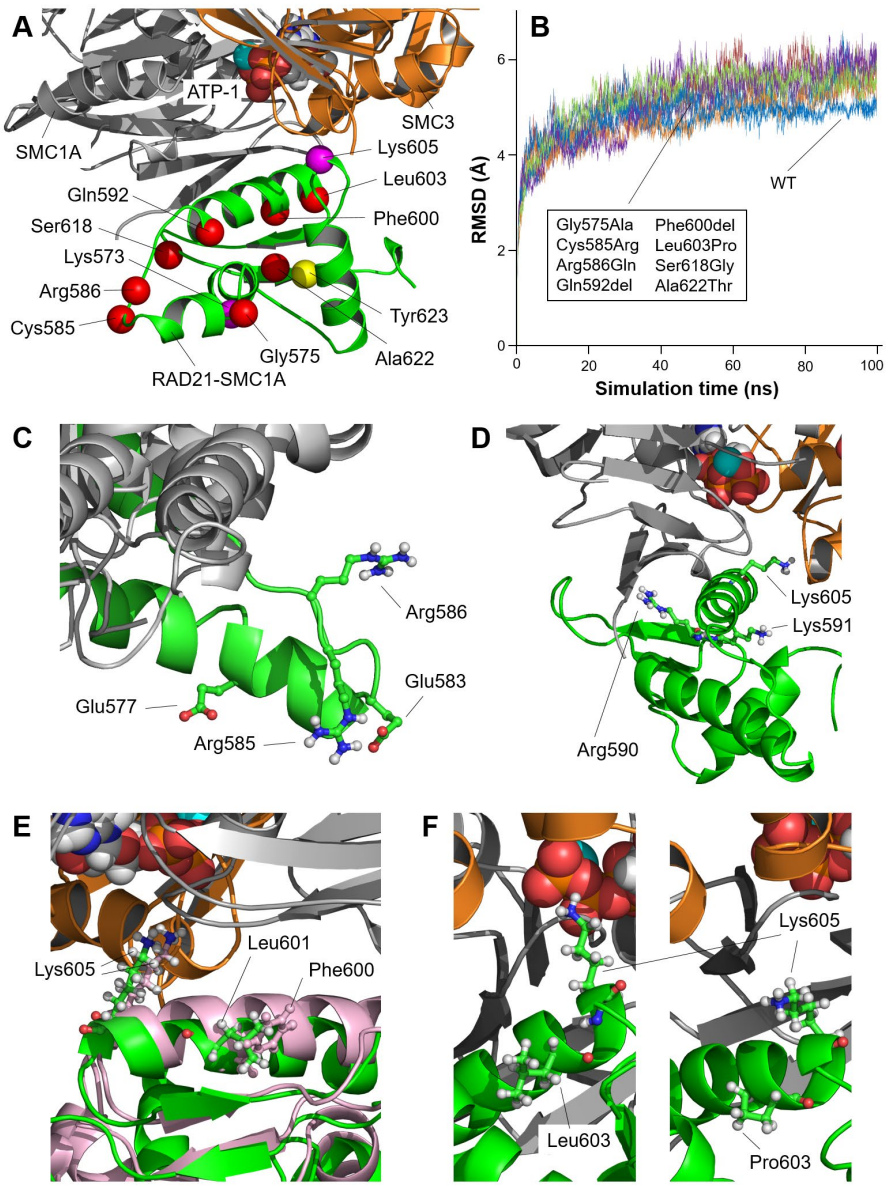
A
ATP-1
SMC3
SMC1A
Lys605
Leu603
Gln592
Phe600
Ser618
Lys573
Arg586
Tyr623
Cys585
Gly575
Ala622
RAD21-SMC1A
B
RMSD (Å)
Simulation time (ns)
WT
Gly575Ala
Cys585Arg
Arg586Gln
Gln592del
Phe600del
Leu603Pro
Ser618Gly
Ala622Thr
C
Arg586
Glu583
Glu577
Arg585
D
Lys605
Lys591
Arg590
E
Leu601
Lys605
Phe600
F
Lys605
Leu603
Pro603

Figure 2 A-F. Structural modeling of RAD21-SMC1A domain bound to the head domain of SMC1A/SMC3 complex.
A. Model for the RAD21-SMC1A domain (residues 543-628, green) associated to the head domains of SMC1A (grey) and SMC3 (orange), close to the ATP molecule in ATPase site 1 (ATP-1) of the SMC1A/SMC3 dimer. Position of mutated residues (Gly575, Cys585, Arg586, Gln592, Phe600, Leu603, Ser6018 and Ala622) is indicated as red spheres. Locations of other important residues (Lys573, Gly575, Lys605, and Thr623) are indicated. Residue Cys585 is located next to residue Arg586. Residue Arg586 interacts through a salt bridge with RAD21 residue Glu577, stabilizing RAD21-SMC1A structure. Three mutated residues (Gln592, Phe600, Leu603) are located in the same alpha helix as key residue Lys605, predicted to maintain the correct positioning of SMC1A-Asn35 at ATPase site 1, putting it into contact with a catalytic water molecule and, thus, allowing progression of the ATPase reaction, pivotal to opening of the cohesin ring and the cyclic process [35]. Variants Ser618Gly and Ala622Thr do not cause structural alterations.
B. Root mean square deviation (RMSD, in Angstroms) of modeled structures (WT, wild-type, blue line; Gly575Ala, red; Cys585Arg, light purple; Arg586Gln, dark green; Gln592del, light blue; Phe600del, orange; Leu603Pro, cyan; Ser618Gly, dark purple; Ala622Thr, light green. No relevant differences in RMSD values demonstrable in the trajectories of the mutated models when compared with wild-type model and with one another.
C. Variant Cys585Arg causes the adjacent Arg586 to lose interaction with Glu577, and both the Arg586 and Glu577 residues change their position in the mutant protein by pointing towards the solvent, which modifies the distribution of charges in the surface of RAD21-SMC1A, while the new Arg585 residue is stabilized in a novel interaction with Glu583.
D. Model for variant Gln592del after 100 ns of MD. New positions of Arg590, Lys591 and Lys605 due to the absence of Gln592 are indicated. Deletion of Gln592 residue causes adjacent Lys591 to be located in the same position as the missing amino acid. This situation promotes a conformational change in the alpha helix, causing the Lys605, which is placed in the same alpha helix, to move away from site 1 of the ATPase.
E. Model for variant Phe600del (green) compared to wild-type model (pink) after 100 ns of MD. Despite the local rearrangement in the alpha helix, distortions of the alpha helix are not relevant as residue Leu601 is placed spatially in the position equivalent to the deleted Phe600 during the MD trajectory, allowing Lys605 to remain in the same position.
F. Structure of wild-type RAD21-SMC1A (left) and variant Leu603Pro (right) models after 100 ns of MD. Presence of mutated Pro603 instead of wild-type Leu603 promotes a local change in the bending angle of the alpha helix in which it is located, resulting in a conformational change in the alpha helix that moves Lys605 out of its initial position close to ATPase site 1.

Figure 3: Clinical phenotype in *RAD21* patients. Anterior-posterior facial views.
F: Family identification number, y: age in years. Family numbers correspond to family numbers in the tables. Ages are indicated below each picture. Intrafamilial variability is illustrated by comparison of facial morphology between the members of family F6 and of family F16. Interfamilial variability is illustrated by comparison of facial morphology between patients F15 and F16a/b who harbor the p.(Arg586*) variant. Pictures of members of F6 and of F17 were republished with permission[24,26]

Table 1. Molecular findings of the presently reported series of individuals with *RAD21* variants. Cohort A: detailed information available, sufficient to score CdLS score; cohort B: insufficient data available to score CdLS score.

PID	Reference	CdLS score†	Exon/intron	Nucleotide change	Predicted amino acid change	Type	Inheritance
COHORT A							
F1	Martinez 2017, (updated)	9	exon 2	c.68G>A	p.(Trp23*)	Nonsense	*de novo*
F2	unpublished	≥7	exon 2	c.194G>A	p.(Arg65Gln)	Missense‡	
F3a	Ansari 2014 P1, (updated)	≥10	intron 3	c.274+1G>A		Splice site	familial (paternal)
F4	Minor 2014 P2, (updated)	12	exon 6	c.592_593dupAG	p.(Ser198Argfs*6)	Frameshift	
F5	unpublished	9	exon 6	c.617_620del	p.(Ile206Thrfs*3)	Frameshift	*de novo*
F6a	Boyle 2017 IV.16, (updated)	12	exon 7	c.704delG	p.(Ser235Ilefs*19)	Frameshift	familial (maternal)
F6b	Boyle 2017 III.1, (updated)	10	exon 7	c.704delG	p.(Ser235Ilefs*19)	Frameshift	familial (parents not tested)
F6c	Boyle 2017 III.2, (updated)	9	exon 7	c.704delG	p.(Ser235Ilefs*19)	Frameshift	familial (parents not tested)
F6d	Boyle 2017 III.5, (updated)	9	exon 7	c.704delG	p.(Ser235Ilefs*19)	Frameshift	familial (parents not tested)
F6e	unpublished	12	exon 7	c.704delG	p.(Ser235Ilefs*19)	Frameshift	familial (maternal)
F7	Dorval 2019	≥11	exon 9	c.943_946del	p.(Glu315Glnfs*9)	Frameshift	*de novo*
F8	Deardorff 2012 P5	≥10	exon 9	c.1127C>G	p.(Pro376Arg)	Missense‡	*de novo*
F9	Kruszka 2019 P14 (updated)	13	exon 10	c.1217_1224del	p.(Lys406Argfs*4)	Frameshift	*de novo*
F10	unpublished	10	exon 11	c.1382C>T	p.(Thr461Ile)	Missense	familial (paternal)
F11a	Minor 2014 P1, (updated)	8	exon 13	c.1621-388_1704+193del	p.(Asp541_Gln568del)	Inframe deletion	familial (maternal)

F11b	Minor 2014 mother P1 (updated)	≥5	exon 13	c.1621-388_1704+193del	p.(Asp541_Gln568del)	665 bp inframe deletion	
F12	unpublished	13	exon 13	c.1635del	p.(Gly547Alafs*65)	Frameshift	*de novo*
F13	Deardorff 2012, P6	≥12	exon 14	c.1753T>C	p.(Cys585Arg)	Missense‡	*de novo*
F14a	unpublished	12	exon 14	c.1753T>C	p.(Cys585Arg)	Missense‡	familial (parents not tested)
F14b	unpublished	≥10	exon 14	c.1753T>C	p.(Cys585Arg)	Missense	familial (parents not tested)
F15	unpublished	≥12	exon 14	c.1756C>T	p.(Arg586*)	Nonsense	
F16a	unpublished	10	exon 14	c.1756C>T	p.(Arg586*)	Nonsense	familial (paternal)
F16b	father, unpublished	≥10	exon 14	c.1756C>T	p.(Arg586*)	Nonsense	
F17	Gudmunsson 2019, updated	8	exon 14	c.1774_1776del	p.(Gln592del)	Inframe deletion‡	*de novo*
F18	unpublished	9	exon 14	c.1800_1802del	p.(Phe600del)	Inframe deletion‡	
F19	Deardorff 2012 P4	≥12	whole gene	arr[hg19] 8q23.3q24.11(116880827-118875305)x1		2 Mb deletion	
F20	unpublished	≥12	whole gene	arr[hg19] 8q23.3q24.11(116915114-119171074)x1		2.3 Mb deletion	*de novo*
F21	Deardorff 2012 P2, McBrein 2008	≥12	whole gene	arr[hg19] 8q23.3q24.12(117571728-119260904)x1		1.7 Mb deletion	*de novo*
F22	unpublished	12	exons 1-9	arr[hg19] 8q24.11(117866471-117893495)x1		27 kb deletion	
COHORT B							
F3b	Ansari 2014, (updated)		intron 3	c.274+1G>A	n/a	Splice site	
F23	Decipher 271431		exon 2	c.16T>G	p.(Phe6Val)	Missense	*de novo*
F24	unpublished		exon 2	c.85delinsCCT	p.(Lys29Profs*10)	Frameshift	
F25a	Decipher 272901		exon 9	c.951del	p.(Ala318Profs*7)	Frameshift	familial (paternal)
F25b	Decipher 272901 father		exon 9	c.951del	p.(Ala318Profs*7)	Frameshift	
F26	Decipher 275402		exon 9	c.1033T>C	p.(Ser345Pro)	Missense‡	*de novo*
F27a	Yuan 2018 P2, (updated)		intron 10	c.1161+1G>A		Splice site	familial (maternal)

F27b	Yuan 2018 mother P2, (updated)		intron 10	c.1161+1G>A		Splice site	
F28a	Kruszka P12/Yuan P1, (updated)		exon 12	c.1550dupC	p.(Glu518Argfs*19)	Frameshift	familial (paternal)
F28b	Kruszka P12 father/Yuan P1 Father, (updated)		exon 12	c.1550dupC	p.(Glu518Argfs*19)	Frameshift	
F29	Lee 2014 P76		exon 14	c.1808T>C	p.(Leu603Pro)	Missense‡	*de novo*
F30a	Bonora 2015 IV.9, (updated)		exon 14	c.[1864G>A];[1864G>A]	p.(Ala622Thr)	Missense‡	familial (both parents)
F30b	Bonora 2015 IV.10, (updated)		exon 14	c.[1864G>A];[1864G>A]	p.(Ala622Thr)	Missense‡	familial (both parents)
F30c	Bonora 2015 IV.11, (updated)		exon 14	c.[1864G>A];[1864G>A]	p.(Ala622Thr)	Missense‡	familial (both parents)
F30d	unpublished		exon 14	c.[1864G>A]	p.(Ala622Thr)	Missense‡	familial (nos)
F30e	unpublished		exon 14	c.[1864G>A]	p.(Ala622Thr)	Missense‡	familial (nos)
F30f	unpublished		exon 14	c.[1864G>A]	p.(Ala622Thr)	Missense‡	familial (nos)
F31	ClinVar		Whole gene	arr[hg19] 8q23.3-24.11(116902507-118942698)x1		2 Mb deletion; Includes several genes	
F32	ClinVar		Whole gene	arr[hg19] 8q23.3-24.11(117509968-118391406)x1		880 kb deletion; includes several genes	
F33	ClinVar		Whole gene	arr[hg19] 8q24.11(117714768-119072307)x1		1.4 Mb deletion; Includes several genes	

†Based on [9]; ≥ defines at least (minor criteria missing) Score <4 is insufficient to indicate molecular testing for CdLS; score 4-8 indicates molecular testing for CdLS indicated; score 9-10 indicates non-classic CdLS; score 11 or higher indicates classic CdLS ; ‡Variants investigated with protein modelling
F: family number; P: patient number in the respective publication; nos: not otherwise specified; VUS, variant of unknown significance

Table 2. Comparison of clinical characteristics of present series of individuals with *RAD21* variants with sufficient clinical data (cohort A) with those in individuals with *SMC1A* and *NIPBL* variants (adapted from [20]).

		RAD21 (n=29)		SMC1A (n= 51)		NIPBL (n=67)	
Clinical characteristics[†]	HPO ID	N pos/N total	percentage	N pos/N total	percentage	N pos/N total	percentage
Sex (Male/Female)		15/14	52/48	14/37	27/73	34/33	51/49
Familial mutation		5/12	42	4/47	9	n/a	n/a
Length at birth <-2SD	HP:0003561	2/18	22	9/31	28	32/43	74
Weight at birth <-2SD	HP:0001511	4/22	18	11/41	27	29/43	67
Prenatal head circumference <-2SD	HP:0000252	7/16	44	8/24	33	39/43	91
Postnatal height <-2SD	HP:0008897	10/27	37	24/38	63	37/43	86
Postnatal weight <-2SD	HP:0004325	3/26	12	14/37	38	39/43	91
Postnatal head circumference <-2SD	HP:0000252	16/28	57	23/36	64	54/62	87
Brachycephaly	HP:0000248	8/19	42	17/42	40	44/67	66
Low anterior/posterior hairline	HP:0000294/HP:0002162	14/23	61	30/43	70	57/67	85
Arched eyebrows	HP:0002553	18/27	67	32/44	73	54/67	81
Synophrys	HP:0000664	19/28	68	37/46	80	61/67	91
Thick eyebrows	HP:0000574	20/24	83	37/46	80	61/67	91
Long eyelashes	HP:0000527	21/26	81	38/45	84	65/67	97
Concave nasal ridge	HP:0011120	24/29	83	20/43	47	57/67	85
Upturned nasal tip	HP:0000463	19/27	70	26/46	57	58/67	87
Short nose	HP:0003196	23/26	88	26/46	57	58/67	87
Long and/or smooth philtrum	HP:0000343/HP:0000319	26/29	90	27/43	63	54/67	81
Thin upper lip vermillion	HP:0000219	23/29	79	33/44	75	22/24	92
Thin lips, downturned corners mouth	HP:0002714	16/27	59	33/46	72	23/24	96

Highly arched palate	HP:0000218	8/22	36	11/37	30	35/67	52
Cleft palate or submucous cleft palate	HP:0000175/HP:0410031	6/25	24	10/45	22	20/67	30
Widely spaced or absent teeth	HP:0000687/HP:0006349	2/20	10	13/44	30	18/23	78
Micrognathia	HP:0000347	8/23	35	18/45	40	50/67	75
Low-set and/or malformed ears	HP:0000369/HP:0000377	14/26	54	18/45	40	45/67	67
Major limb malformation	HP:0001180/HP:0009776	0/29	0	0/49	0	17/67	25
Small hands	HP:0200055	5/27	19	32/45	71	53/63	84
Proximally placed thumb	HP:0009623	6/18	33	18/44	41	11/20	55
Clinodactyly 5th finger	HP:0004209	13/24	54	21/45	47	42/63	67
Short 5th finger	HP:0009237	23/28	82	21/45	47	42/63	67
Syndactyly hands	HP:0006101	1/19	5	1/37	3	4/63	6
Abnormal palmar crease	HP:0010490	9/21	43	5/40	13	21/29	72
Dislocated elbow/abnormal extension	HP:0005021/HP:0001377	11/24	46	2/40	5	20/34	59
Small feet	HP:0001773	3/27	11	29/44	66	65/67	97
Syndactyly 2nd-3rd toes	HP:0004691	4/24	17	13/46	28	21/66	32
Scoliosis	HP:0002650	2/20	10	4/40	10	1/42	2
Hip dislocation or dysplasia	HP:0002827/HP:0001385	2/19	11	2/40	5		
Ptosis	HP:0000508	11/26	42	4/40	10	8/42	19
Visual impairment	HP:0000505	0/24	0	20/38	53	29/66	44
Myopia ≥ -6.00 D	HP:0011003	≤2/24‡	≤8	11/40	28	6/40	15
Hearing loss	HP:0000365	8/24	33	16/39	41	43/66	65
Seizures	HP:0001250	2/22	9	20/44	45	10/66	15
Cutis marmorata	HP:0000965	3/23	13	19/44	43	27/43	63
Hirsutism	HP:0001007	10/26	38	37/47	79	37/43	86

CNS major and minor malformations (MRI Brain)	HP:0012443	2/5	40	5/43	12		
Heart (major and minor)	HP:0001627	9/23	39	13/44	30	18/66	27
Major malformation of gut	HP:0012718	4/30	13	3/44	7	6/24	25
Diaphragmatic hernia	HP:0000776	1/30	3	1/40	3	6/24	25
Gastroesophageal reflux disease	HP:0002020	13/25	52	25/42	60	47/66	71
Genitourinary system major[§]	HP:0000119	1/20	5	4/42	10	0/67	0
Genitourinary system minor	HP:0000119	8/23	35	9/40	23	46/67	69

HPO ID, Human phenotype ontology identifier; CNS, central nervous system; † Only features which could be compared across at least two cohorts are presented. Full clinical description with individual data are presented in supplementary Table S3; ‡2 others among the 24cases have myopia but unspecified severity; § Uni/bilateral renal anomalies.

Table 3. Cognitive and behavioral characteristics of individuals with *RAD21* variants with sufficient clinical data (cohort A) with those in individuals with *SMC1A* and *NIPBL* variants (adapted from [20-22]).

	RAD21 (n=29)		SMC1A (n= 51)		NIPBL (n=67)	
	N pos/ N total	%	N pos/ N total	%	N pos/ N total	%
Cognitive functioning†						
Normal cognition	3/29‡	10	3/28	11	0/58	0
Mild disability (HP:0001256)	13/29	45	6/28	21	4/58	7
Moderate disability (HP:0002342)	4/29§	14	9/28	32	16/58	28
Severe disability (HP:0010864)	0/29	0	6/28	21	27/58	47
Profound disability (HP:0002187)	0/29	0	4/28	14	11/58	19
Disability present, severity unspecified (HP:0001249)	2/29	7				
Developmental problems, too young to determine reliably cognitive functioning (HP:0012759)	7/29	24				
Developmental milestones¶						
Sitting without support		100††		75††		54††
Attained on target (age <12 months)	10/10		n/a		n/a	
Attained before age 3 years	10/10		18/24		28/52	
Attained later			3/24		23/52	
Not attained yet (in patients aged ≥5 years)			3/24		1/52	
First words		100††		35††		8††
Attained on target (age <15 months)	6/15		n/a		n/a	
Attained before age 3 years	15/15		7/20		4/53	
Attained later			4/20		16/53	
Not attained yet patients aged ≥5 years)			9/20		33/53	
Walking without support		100††		57††		≥29††,§§
Attained on target (age <18 months)	12/16		n/a		1/52	
Attained before age 3 years	16/16		17/30		2/52	
Attained later			9/30		12/52	
Not attained yet (in patients aged ≥5 years)			4/30		19/52	
Delay on one or more milestone	12/16	75	18/20	90	51/52	98
Behavior‡‡						
Attention deficit disorder +/- hyperactivity	8/23	35				
Obsessive-compulsive behavior	6/19	32	10/26¶¶	38		
Anxiety	10/19	53				
Constant roaming	3/15	20				
Aggression	1/16	6			12/15¶¶	80
Self-injurious behavior	1/18	6	11/31	35	47/61	77
Extreme shyness or withdrawal	0/17	0				
Autistic-like features	7/20	35	18/31¶¶	56	9/13¶¶	69
One or more behavioral domains affected	14/25	56				

HP, Human phenotype ontology identifier; †RAD21, 8 formal test results, others physician reported data. Equivalent HP is shown between brackets; ‡ includes 2 adults with learning disabilities but reported normal cognitive functioning; § including 2 moderate/severe; ¶Only scored if child was older than target age; ††percentage of individuals that attain the milestone before age 3 years; ‡‡RAD21: 5 formal test results, others physician reported data; §§ Including 18 patients that attained the milestone late, but age unknown; ¶¶based on formal testing.