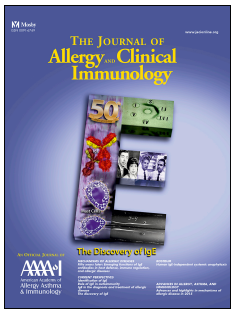
# Accepted Manuscript




Reduced immunoglobulin gene diversity in patients with Cornelia de Lange syndrome

Andrea Björkman, PhD, Likun Du, PhD, Mirjam van der Burg, PhD, Valerie Cormier-Daire, MD, Guntram Borck, PhD, Juan Pié, MD, Britt-Marie Anderlid, MD, Lennart Hammarström, MD, PhD, Lena Ström, PhD, Jean-Pierre de Villartay, PhD, David Kipling, PhD, Deborah Dunn Walters, PhD, Qiang Pan-Hammarström, MD, PhD

PII: S0091-6749(17)31266-6

DOI: 10.1016/j.jaci.2017.06.043

Reference: YMAI 12947

To appear in: *Journal of Allergy and Clinical Immunology*

Received Date: 30 August 2016

Revised Date: 12 June 2017

Accepted Date: 20 June 2017

Please cite this article as: Björkman A, Du L, van der Burg M, Cormier-Daire V, Borck G, Pié J, Anderlid B-M, Hammarström L, Ström L, Villartay J-Pd, Kipling D, Walters DD, Pan-Hammarström Q, Reduced immunoglobulin gene diversity in patients with Cornelia de Lange syndrome, *Journal of Allergy and Clinical Immunology* (2017), doi: 10.1016/j.jaci.2017.06.043.

# Reduced immunoglobulin gene diversity in patients with Cornelia de Lange syndrome

Andrea Björkman, PhD[1#], Likun Du, PhD[1#], Mirjam van der Burg, PhD[2], Valerie Cormier-Daire, MD[3], Guntram Borck, PhD[4], Juan Pié, MD[5], Britt-Marie Anderlid, MD[6], Lennart Hammarström, MD, PhD[1], Lena Ström, PhD[7], Jean-Pierre de Villartay, PhD[8], David Kipling, PhD[9], Deborah Dunn Walters, PhD[10,11], Qiang Pan-Hammarström, MD, PhD[1*]

1. Department of Laboratory Medicine, Karolinska Institutet, 171 77 Stockholm, Sweden
2. Department of Immunology, Erasmus MC, Rotterdam, 3015 CN, The Netherlands
3. Department of Genetics, INSERM U781, Hospital Necker, 75743 Paris, France
4. Institute of Human Genetics, University of Ulm, 89081 Ulm, Germany
5. Unit of Clinical Genetics and Functional Genomics, Departments of Pharmacology-Physiology and Pediatrics, School of Medicine, University of Zaragoza, E-50009 Zaragoza, Spain
6. Department of Clinical Genetics, Karolinska University Hospital, 171 76 Stockholm, Sweden
7. Department of Cell and Molecular Biology, Karolinska Institutet, 171 77 Stockholm, Sweden
8. Université Paris-Descartes, Faculté de Médicine René Descartes, Site Necker, Institut Fédératif de Recherche, F-71015 Paris, France
9. Division of Cancer and Genetics, School of Medicine, Cardiff University, Cardiff CF14 4XN, UK
10. Department of Immunobiology, King's College London School of Medicine, SE1 9RT London, UK
11. Faculty of Health & Medical Sciences, University of Surrey, Guildford GU2 7XH, UK

# These authors contributed equally to this study.

*Corresponding author: Prof. Qiang Pan-Hammarström (qiang.pan-hammarstrom@ki.se), Div. of Clinical Immunology, F79, Department of Laboratory medicine, SE141 86 Stockholm, Sweden, Phone; + 46 8 52483592, Fax; + 46 8 52483588

This work was supported by the European Research Council, the Swedish Research Council, the Swedish Cancer Society, the Center for Innovative Medicine at Karolinska Institutet (CIMED) and the Swedish Childhood Cancer Foundation.

**Key messages:**

- B-cells from CdLS patients have altered *IGH* repertoires with a reduced diversity and skewed *VH* usage.
- The frequency of somatic hypermutations is reduced in the *IGHV* regions in B-cells from *NIPBL* mutated CdLS patients.
- Inefficient immunoglobulin gene recombination/diversification may underlie the susceptibility to infection in CdLS patients.

**Capsule summary:** Altered *IGH* repertoires with reduced diversity, skewed *VH* gene usage and reduced frequency of somatic hypermutation were observed in CdLS patients.

**Key words:** CdLS, cohesin, NIPBL, SMC1A, immunoglobulin, V(D)J recombination, somatic hypermutation, B-cells, human.

***To the Editor,***

B-cells rely on a broad receptor repertoire to provide protection against a wide range of pathogens. This is in part achieved via V(D)J recombination which, by assembling various combinations of variable (V), diversity (D) and joining (J) genes, creates different immunoglobulin (Ig) V regions(1). The recombination processes is initiated by RAG1/RAG2 enzymes and requires a functional non-homologous end joining (NHEJ) machinery. B-cells can further diversify their Ig V regions through somatic hypermutation (SHM), to improve the affinity between the antibody and antigen, and switch the isotype of antibody produced by class switch recombination (CSR). Both processes are initiated by activation-induced cytidine deaminase (AID) and rely on transcription and a number of DNA repair mechanisms.

Cornelia de Lange syndrome (CdLS) is a rare, multisystem developmental disorder characterized by typical facial features, intellectual disability and multiple congenital anomalies(2). Most CdLS patients have deleterious mutations in the gene encoding the cohesin loader NIPBL, but mutations in other cohesin related genes *SMC1A*, *SMC3*, *PDS5B*, *RAD21* or *HDAC8* have also been identified in selected patients. Cohesin has been implicated in regulation of sister chromatid cohesion, transcription, long-range gene interactions and DNA repair(3).

Aberrant CSR patterns have been observed in *NIPBL* mutated patient's B-cells(4). Here we further investigated whether V(D)J recombination or SHM is affected in CdLS patients. In chromosomal-integrated V(D)J reporter assays, a reduced substrate

recombination efficiency was observed in *NIPBL* knockdown human fibroblast- or murine pro-B-cells (Fig. E1). We next analyzed the *IGH* repertoire in five *NIPBL* and three *SMC1A* mutated CdLS patients (Table E1). The overall diversity of the *IGHV* repertoire, estimated by either the proportion of sequences with a unique CDR3 divided by the total number of sequences, or by D50, which is a measurement of the evenness of the distribution of B cell clones, was significantly lower in the patients with *NIPBL* or *SMC1A* mutations (Fig. 1a, b). Thus, the CdLS patients had a reduced overall diversity of their *IGHV* regions, with an overrepresentation of large B cell clones.

Furthermore, in *NIPBL* mutated patients with the classical form disease and in *SMC1A* mutated patients, the frequency of *IGHV* genes located in the most proximal one third of the *IGH* locus (about 250 kb) was increased (Fig. E2a, b, Fig. 1c). The observed skewed pattern of *VH* genes in the patients is likely to be a result of B-cell intrinsic changes, as only sequences resulting from unproductive rearrangement (successfully rearranged, but out of reading-frame or containing stop codons) were included in this analysis and an influence of antigen selection can thus be excluded.

The overall mutation frequency in the unproductive, most commonly mutated *IGHV* genes was also reduced in the CdLS patients, reaching a statistical significance for the *NIPBL* mutated group (Fig. 2a). The proportion of un-mutated sequences was increased in the *NIPBL* mutated patients, especially in those with the classical form of disease (Fig. 2a). The pattern of base pair substitution in the V regions was, however, largely normal in the patients (Fig. 2b).

Our data suggest that the cohesion associated proteins NIPBL and SMC1A are involved in the regulation of V(D)J recombination in human B-cells, possibly through regulation of locus contraction. Increased usage of proximally located *IghV* genes has previously been observed in mice deficient in factors associated with locus contraction, such as Ikaros, Yin Yang 1 and Pax5 and the cohesin interaction partner CCCTC binding factor has been suggested to regulate locus contraction and V(D)J recombination at the mouse *Igh* locus(5, 6), (7).

Another possible explanation of how NIPBL/SMC1A could affect V(D)J recombination could be a change in regulation of DNA repair. We have previously observed a correlation between heterozygous *NIPBL* loss-of-function mutations and increased sensitivity to γ-radiation and a shift towards the use of a microhomology-based end joining during CSR, suggesting that NIPBL regulates the NHEJ process(4). However, the V(D)J coding junctions generated *in vivo* in CdLS patients showed a normal repair pattern (Table E2). Furthermore, the expression of a number of key V(D)J recombination factors was largely normal in *NIPBL* knockdown cells (Fig. E3b). Thus, the core NHEJ machinery appears to be retained in CdLS patients.

We also found a reduced frequency of mutations within the *IGHV* regions in CdLS patients. NIPBL and SMC1A could both be involved in the SHM process through regulation of transcription or RNA polymerase pausing, which promotes the formation of single-stranded DNA and AID targeting. In support of this notion, there was a reduced number of mutations observed in the RGYW motifs (containing the

AID hotspot in the bottom strand) in *NIPBL*$^{+/-}$ and *SMC1A*$^{-}$ cells, suggesting an impaired mutagenesis of C residues on the bottom strand (Fig. 2c). We have previously observed the opposite pattern in NBS1-deficient cells[9]. Thus, the targeting of AID and/or linked repair seem to be asymmetric during SHM and this can be regulated by a number of factors including the cohesin associated factors.

CdLS is not traditionally considered as an immunodeficiency disorder, but these patients have an increased prevalence of severe infections[10]. Immunodeficiency, including mild B-cell lymphopenia, reduced switched memory B cells and/ reduced serum Ig levels has indeed been observed in previous case reports[10] and in our own patients (Table E3). Complete knockout of NIPBL is incompatible with survival and heterozygous nonsense mutation or frame shift deletions in *NIPBL* are usually associated with a more severe clinical phenotype than those with missense mutations and in frame deletions (classical vs. mild form), suggesting a dose dependent role of the cohesin associated proteins. The *SMC1A* mutations are missense or in frame deletions and are generally associated with a milder CdLS phenotype. Thus, mutations identified in CdLS patients can be considered as hypomorphic and/or result in haploinsufficiency, and a more significant alteration in Ig gene diversifications and a more severe immunodeficiency would be expected in the event of a complete loss of these cohesin associated proteins.

In summary, the less diversified *VH* genes due a reduced efficiency of V(D)J recombination, together with an inefficient CSR efficiency, might contribute to the frequent infections in CdLS patients.

Andrea Björkman, PhD[1#]

Likun Du, PhD[1#]

Miriam van der Burg, PhD[2]

Valerie Cormier-Daire, MD[3]

Guntram Borck, PhD[4]

Juan Pié, MD[5]

Britt-Marie Anderlid, MD[6]

Lennart Hammarström, MD, PhD[1]

Lena Ström, PhD[7]

Jean-Pierre de Villartay, PhD[8]

David Kipling, PhD[9]

Deborah Dunn Walters, PhD[10,11]

Qiang Pan-Hammarström, MD, PhD[1]

1. Department of Laboratory Medicine, Karolinska Institutet, 171 77 Stockholm, Sweden

2. Department of Immunology, Erasmus MC, Rotterdam, 3015 CN, The Netherlands

3. Department of Genetics, INSERM U781, Hospital Necker, 75743 Paris, France

4. Institute of Human Genetics, University of Ulm, 89081 Ulm, Germany

5. Unit of Clinical Genetics and Functional Genomics, Departments of Pharmacology-Physiology and Pediatrics, School of Medicine, University of Zaragoza, E-50009 Zaragoza, Spain

6. Department of Clinical Genetics, Karolinska University Hospital, 171 76 Stockholm, Sweden

7. Department of Cell and Molecular Biology, Karolinska Institutet, 171 77 Stockholm, Sweden

8. Université Paris-Descartes, Faculté de Médicine René Descartes, Site Necker, Institut Fédératif de Recherche, F-71015 Paris, France

9. Division of Cancer and Genetics, School of Medicine, Cardiff University, Cardiff CF14 4XN, UK

10. Department of Immunobiology, King's College London School of Medicine, SE1 9RT London, UK

11. Faculty of Health & Medical Sciences, University of Surrey, Guildford GU2 7XH, UK

## References

1. Schatz DG, Ji Y. Recombination centres and the orchestration of V(D)J recombination. Nat Rev Immunol. 2011;11(4):251-63.
2. Liu J, Baynam G. Cornelia de Lange syndrome. Adv Exp Med Biol. 2010;685:111-23.
3. Mehta GD, Kumar R, Srivastava S, Ghosh SK. Cohesin: functions beyond sister chromatid cohesion. FEBS Lett. 2013;587(15):2299-312.
4. Enervald E, Du L, Visnes T, Bjorkman A, Lindgren E, Wincent J, et al. A regulatory role for the cohesin loader NIPBL in nonhomologous end joining during immunoglobulin class switch recombination. J Exp Med. 2013.
5. Degner SC, Verma-Gaur J, Wong TP, Bossen C, Iverson GM, Torkamani A, et al. CCCTC-binding factor (CTCF) and cohesin influence the genomic architecture of the Igh locus and antisense transcription in pro-B cells. Proc Natl Acad Sci U S A. 2011;108(23):9566-71.
6. Degner SC, Wong TP, Jankevicius G, Feeney AJ. Cutting edge: developmental stage-specific recruitment of cohesin to CTCF sites throughout immunoglobulin loci during B lymphocyte development. J Immunol. 2009;182(1):44-8.
7. Guo C, Yoon HS, Franklin A, Jain S, Ebert A, Cheng HL, et al. CTCF-binding elements mediate control of V(D)J recombination. Nature. 2011;477(7365):424-30.

**Figure legends**

**Fig. 1** ***Reduced IGH diversities and increased usage of proximal IGHV genes in CdLS patients***

a) Diversities (unique/total CDR3s) and b) the proportion of unique B-cell clones that account for the cumulative 50 percent of the total CDR3s in the samples, D50, are shown. Significant differences compared to controls are indicated by stars, $**p<0.01$, $***p<0.001$, Student's t-test. Each dot represents one individual. Black or white squares indicate *NIPBL* mutated patients with the classical or mild form of disease, respectively. c) Average frequencies of proximally, middle and distally located *V* genes (in relation to *D* and *J* genes) in CdLS patients. Schematic picture (not to scale) of the *IGH* locus on top. Proximal, middle or distal V gene group each encompasses approximately 250 kb of the *IGHV* locus and contains 15, 20 and 19 functional *V* genes respectively. The number of sequences used in the analysis is indicated below each group. The proportion of proximal, middle or distal *V* genes for each individual was regarded as one data point, and the patient and control groups were then compared by Student t-test. Significant differences compared to controls are indicated by arrows. ↑ and ↓, increased or decreased compared to controls, respectively.

**Fig. 2** ***Mutation frequency and substitution pattern in IGHV regions from NIPBL$^{+/-}$ and SMC1A$^{-}$ B cells*** a) Mutation frequencies are shown on the left. Each dot represents one individual. Black or white squares indicate *NIPBL*$^{+/-}$ patients with a classical or mild form of disease, respectively. The percentages of sequences with 0, 1-5, 6-10 or >10 mutations in the IGHV regions from control, *NIPBL*$^{+/-}$ or *SMC1A*$^{-}$

cells are depicted in pie charts to the right. *IGHV1-18*, *IGHV1-69*, *IGHV3-23*, *IGHV3-30*, *IGHV3-33* and *IGHV4-34* sequences were included in the analysis. Significant differences as compared to controls are indicated by star(s), *$p$<0.05; Student's t-test. b) Frequencies of substitutions in the *IGHV* genes. $\chi^2$ test, Bonferroni corrected. c) Frequency of mutations within W(A/T)R(G/A)CY(C/T)/RGYW motifs in the *IGHV* genes analyzed in a). Stars indicate significant differences between *NIPBL*$^{+/-}$ (black star), *SMC1A*$^{-}$ (grey star) compared to control cells, *$p$<0.05; $\chi^2$ test, Bonferroni corrected.

**Fig. 1**

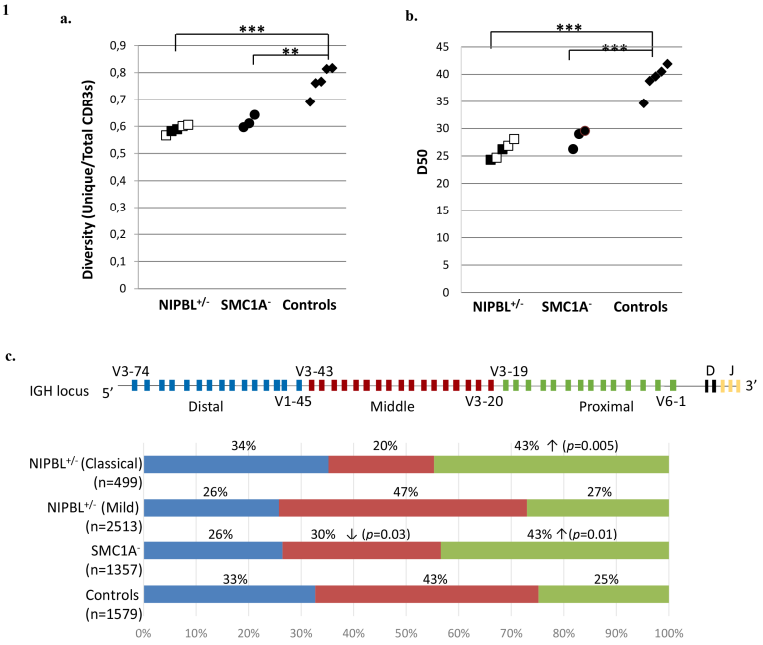
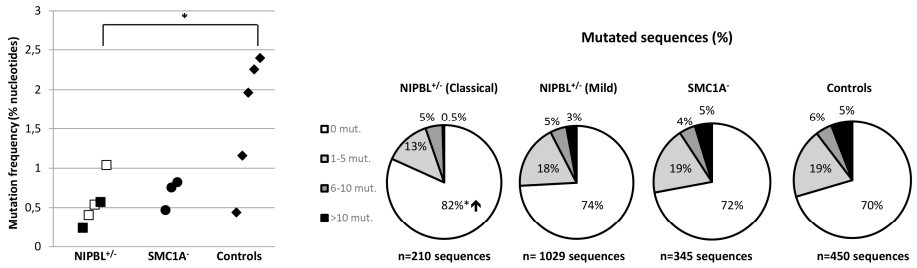


**Fig. 2**

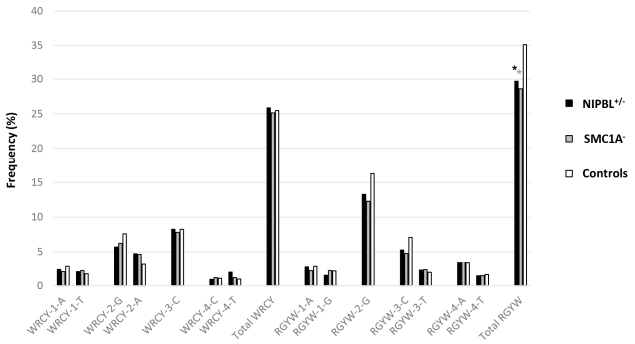
**a.**




**b.**

| | | NIPBL$^{+/-}$ (%) (1346 mutations) | | | | | | SMC1A$^{-}$ (%) (690 mutations) | | | | | | Controls (%) (918) mutations) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **To** | A | C | G | T | Total | | A | C | G | T | Total | | A | C | G | T | Total |
| **From** | A | / | 5.3 | 13.8 | 7.8 | 26.9 | A | / | 6.1 | 12.8 | 6.8 | 25.7 | A | / | 6.5 | 11.8 | 6.4 | 24.7 |
| | C | 4.2 | / | 7.3 | 12.8 | 24.3 | C | 3.8 | / | 7.5 | 11.6 | 22.9 | C | 4.1 | / | 6.5 | 13.9 | 24.6 |
| | G | 17.2 | 10.0 | / | 4.5 | 31.7 | G | 17.8 | 10.4 | / | 4.2 | 32.4 | G | 18.4 | 12.5 | / | 5.6 | 36.9 |
| | T | 5.2 | 7.0 | 4.9 | / | 17.1 | T | 5.2 | 9.0 | 4.8 | / | 19.0 | T | 2.9 | 8.3 | 2.8 | / | 13.9 |
| | Total | 26.6 | 22.3 | 26.0 | 25.1 | 100 | Total | 26.8 | 25.5 | 25.1 | 22.6 | 100 | Total | 25.5 | 27.3 | 21.1 | 25.9 | 100 |
| | Mutations at G or C site | | | | | 56.0 | Mutations at G or C site | | | | | 55.3 | Mutations at G or C site | | | | | 61.4 |
| | Transitions | | | | | 50.8 | Transitions | | | | | 51.2 | Transitions | | | | | 47.1 |

**c.**



# Supplemental Material and Methods

## *Patients*

Eight CdLS patients were included in the study (Table E1). Five had heterozygous mutations in the *NIPBL* gene, of which two (P3, P6) had a more severe, classical form of CdLS, whereas three (P8, P10, P11) had a milder form of disease. In addition, three male patients had mutations within the *SMC1A* gene, located on the X chromosome (SMC1A-1, SMC1A-2 and SMC1A-4). The clinical features of seven of the eight patients have been described previously[1-3]. Seven unaffected controls were also included in the study (Table E1). DNA was isolated from peripheral blood using standard methods. The study was approved by the institutional review board at the Karolinska Institutet.

## *V(D)J recombination substrate assays for assessment of recombination efficiency*

A retroviral vector pMX-RSS-GFP/IRES-hCD4 (MX-RSS-INV), which contains an anti-sense-orientated RSS-flanked EGFP reporter and an internal ribosomal entry site (IRES)-linked human CD4 cassette[6] was transduced into HEK293FT cells together with three viral packaging vectors (pGag-Pol, pRev and pVSV-G) by using lipofectamine 2000 (Thermo Fisher Scientific, Waltham, MA, USA). 48 hours later, the medium containing virus was collected and used to infect the human fibroblast cell line from an unaffected control. 72 hours later the transduced cells were sorted by flow-cytometry using human CD4 antibody (Biolegend, San Diego, CA, USA). The $CD4^+$ cells were expanded in the culture for 4 days and then seeded into 6-well plate ($10^5$ cells per well) and treated with either NIPBL or control siRNA (GE Dharmacon, Lafayete, CO, USA)[4] for 48 hours. The cells were subsequently transfected with 0.9 μg of RAG1 and 0.75 μg RAG2 plasmids using Turbofect (Fermentas, Waltham, MA, USA). 48 hours after transfection, cells were collected and recombination activity were assessed as percentages of $GFP^+$ cells by flow-cytometry. This chromosomal integrated reporter assay allows the assessment of V(D)J recombination and infection efficiency by measurement of expression of GFP or CD4 respectively. The gene knockdown efficiency was monitored by real-time PCR (Fig. E1b) as described previously[4].

A murine B-abl pro-B-cell line (with a *Bcl2* transgene) carrying the chromosome integrated MX-RSS-INV substrate was previously described[7]. The cells were treated with either *Nipbl* or control smart pool mouse siRNA (GE Dharmacon, Lafayette, CO, USA) using a nucleofector kit (Lonza, Basel, Switzerland) for 48 hours and then treated with 0.3 μM abl kinase inhibitor

(PD-180970, Sigma-Aldrich, St. Louis, MO, USA), which leads to G1 cell cycle arrest and an induction/stabilization of expression of RAG1/RAG2[8]. Recombination activity were subsequently assessed as percentages of $GFP^+$ cells by flow-cytometry. The gene knockdown efficiency was monitored by real-time PCR (Fig. E1c).

***Amplification and high throughput sequencing of IGHV regions from CdLS patients***

Ig genes were amplified using a semi-nested PCR as previously described[9] with the exception that 100 ng of genomic DNA and 250 nM of each downstream *IGHJ* primer were used in the PCR reaction. Sequencing was performed on the Roche 454 GS FLX Titanium Sequencer (Agowa GmbH, Germany). V-QUEST[10,11], allowing detection of indels, was used to determine the amino acid sequence of the complementary determining region 3 (CDR3), as well as the *V*-, *D*- and *J*-genes used. Sequences not spanning the V(D)J-junction were removed and the remaining sequences were used in subsequent analyses. Any large deletions that resulted in missing entire CDR3 would therefore not be recorded. The D50-value was calculated by dividing X with Y and multiplying with 100, where X= number of unique CDR3s that account for the cumulative 50 percent of the total sequences, Y= number of unique CDR3s. A sample with an equal distribution of all clones would have a D50 of 50, whereas a number less than 50 indicates the presence of uneven, larger clones[12]. Chimeric sequences, which are artifacts created during the PCR reaction that contains more than one *IGHV* region, were detected by visual inspection. If the sequence matched more than one *IGHV* gene, at either end, the sequence was omitted from *IGHV* gene usage analysis. For SHM analysis, chimeric sequences were also omitted, except if the sequence was chimeric at the first half of the *IGHV* region sequence, then the remaining part was still included in the analysis. Only unique, unproductive sequences (*VH* and *JH* genes are not in the same reading frame or containing premature stop codons) were included for the *VH* gene usage and SHM analyses, in order to avoid any antigen selection bias from the expressed repertoire. The error rate of 454 sequencing was previously determined to be less than 1 error per 1300 bp when indels were excluded[9].

**Real-time PCR**

One µg of total RNA was used to synthesize cDNA according to the manufacturer's protocol (cDNA synthesis kit, GE Healthcare). Expression of human *XRCC6*, *LIG4*, *DCLRE1C*, *PRKDC* or *NIPBL* in the fibroblast cell line treated with control or *NIPBL* siRNA was determined by quantitative real-time PCR assays with a similar setting as described

previously[4]. Primer sequences are available upon request. Briefly, the real-time PCR amplification was performed using the Kapa SYBR Fast qPCR Master Mix kit (Kapa Biosystems). *ACTB* was used as housekeeping gene for calculation of relative expression levels. Expression of *Nipbl* in the murine B-abl pro-B-cell line treated with control or *Nipbl* siRNA was determined by a similar strategy, with mouse *Gapdh* as housekeeping gene.

**References**


1. Pie J, Gil-Rodriguez MC, Ciero M, Lopez-Vinas E, Ribate MP, Arnedo M, et al. Mutations and variants in the cohesion factor genes NIPBL, SMC1A, and SMC3 in a cohort of 30 unrelated patients with Cornelia de Lange syndrome. Am J Med Genet A 2010; 152A:924-9.
2. Schoumans J, Wincent J, Barbaro M, Djureinovic T, Maguire P, Forsberg L, et al. Comprehensive mutational analysis of a cohort of Swedish Cornelia de Lange syndrome patients. Eur J Hum Genet 2007; 15:143-9.
3. Borck G, Zarhrate M, Bonnefont JP, Munnich A, Cormier-Daire V, Colleaux L. Incidence and clinical features of X-linked Cornelia de Lange syndrome due to SMC1L1 mutations. Hum Mutat 2007; 28:205-6.
4. Enervald E, Du L, Visnes T, Bjorkman A, Lindgren E, Wincent J, et al. A regulatory role for the cohesin loader NIPBL in nonhomologous end joining during immunoglobulin class switch recombination. J Exp Med 2013; 210(12):2503-13
5. Nicolas N, Moshous D, Cavazzana-Calvo M, Papadopoulo D, de Chasseval R, Le Deist F, et al. A human severe combined immunodeficiency (SCID) condition with increased sensitivity to ionizing radiations and impaired V(D)J rearrangements defines a new DNA recombination/repair deficiency. J Exp Med 1998; 188:627-34.
6. Liang HE, Hsu LY, Cado D, Cowell LG, Kelsoe G, Schlissel MS. The "dispensable" portion of RAG2 is necessary for efficient V-to-DJ rearrangement during B and T cell development. Immunity, 2002, 17(5): 639-51.
7. Lescale C, Abramowski V, Bedora-Faure M, Murigneux V, Vera G, Roth BD, et al. RAG2 and XLF/Cernunnos interplay reveals a novel role for the RAG complex in DNA repair. Nat. Commun 2016; 2;7:10529
8. Bredemeyer AL. Sharma G., Huang C, Helmink BA, Walker LM, Khor K.C., et al. ATM stabilizes DNA double-strand-break complexes during V(D)J recombination. Nature 2006; 442:466–70

9. Wu YC, Kipling D, Leong HS, Martin V, Ademokun AA, Dunn-Walters DK. High-throughput immunoglobulin repertoire analysis distinguishes between human IgM memory and switched memory B-cell populations. Blood 2010; 116:1070-8.
10. Giudicelli V, Brochet X, Lefranc MP. IMGT/V-QUEST: IMGT standardized analysis of the immunoglobulin (IG) and T cell receptor (TR) nucleotide sequences. Cold Spring Harb Protoc 2011; 2011:695-715.
11. Brochet X, Lefranc MP, Giudicelli V. IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis. Nucleic Acids Res 2008; 36:W503-8.
12. Yang Y, Wang C, Yang Q, Kantor AB, Chu H, Ghosn EE, et al. Distinct mechanisms define murine B cell lineage immunoglobulin heavy chain (IgH) repertoires. Elife 2015; 4:e09083.

**Table E1**: Patients included in study

| Patient | Age; Sex | Mutation | Diagnosis | Total seq. | Productive/ Unproductive ratio[a] | Seq. with unique CDR3 | Diversity[b] | Unproductive seq. with unique CDR3, not chimeric[c] | Reference |
|---|---|---|---|---|---|---|---|---|---|
| P3 | 17y; F | NIPBL$^{+/-}$, c.6250_6255del, p.V2084-V2085, deletion | Classical CdLS | 806 | 0.69/0.31 | 475 | 0.59 | 125 | [2, 4] |
| P6 | 2y; M | NIPBL$^{+/-}$, c.4593T>A, p.Y1531X, nonsense | Classical CdLS | 1874 | 0.68/0.32 | 1091 | 0.58 | 374 | [2, 4] |
| P8 | 4y; F | NIPBL$^{+/-}$, c.4321G>T, p.V1441L, splice site[d] | Mild CdLS | 11776 | 0.66/0.34 | 7070 | 0.60 | 1938 | [1, 4] |
| P10 | 10y; M | NIPBL$^{+/-}$, c.6242G>C, p.G2081A, missense | Mild CdLS | 2261 | 0.66/0.34 | 1386 | 0.61 | 360 | [1, 4] |
| P11 | 14y; M | NIPBL$^{+/-}$, c.8387A>G, p.Y2796C, missense | Mild CdLS | 1191 | 0.64/0.36 | 675 | 0.57 | 215 | This report |
| SMC1A-1 | 2y; M | SMC1A$^{-}$, c.587G>A, p.R196H, missense | CdLS | 2497 | 0.66/0.34 | 1489 | 0.60 | 405 | [3, 4] |
| SMC1A-2 | 5y; M | SMC1A$^{-}$, c.3254A>G, p.Y1085C, missense | CdLS | 3308 | 0.68/0.32 | 2022 | 0.61 | 550 | [3, 4] |
| SMC1A-4 | 9y; M | SMC1A$^{-}$, c.3568A>G, p.K1190E, missense | CdLS | 2163 | 0.66/0.34 | 1390 | 0.64 | 402 | [4] |
| C19 | 19y; M | - | Unaffected | 276 | 0.62/0.38 | 210 | 0.76 | 41 | This report |
| C31 | 19y; ? | - | Unaffected | 316 | 0.62/0.38 | 257 | 0.81 | 71 | This report |
| C34 | 13y; ? | - | Unaffected | 660 | 0.65/0.35 | 506 | 0.77 | 113 | This report |
| C54 | 13y; ? | - | Unaffected | 1840 | 0.63/0.37 | 1272 | 0.69 | 321 | This report |
| gCont[e] | 2y, 5y, 7y; ? | - | Unaffected | 3482 | 0.53/0.47 | 2845 | 0.82 | 1033 | This report |

a. Productive: as a result of V(D)J recombination, the VH, DH and JH genes are in the correct translational reading-frame.
b. Diversity was calculated by dividing sequences with unique CDR3s with the total sequences. Sequences not spanning the CDR3 region were omitted from calculation.
c. Only unproductive rearrangement (out of reading-frame or contain a stop codon), unique sequences, not chimeric at both ends, were included in V gene analysis.
d. This mutation has been shown to yield two transcripts, of which one bears an exon 20 deletion and the other is of ordinary size but with a V1441L change[1].
e. gCont contained a pool of DNA from 3 individuals. y: years; F: female; M: male; seq.: sequence.

**Table E2:** Characterization[a] of *in vivo* generated V(D)J coding junctions[b]

| Study groups | Average CDR3 lengths (bp) | Average no. of P nts (bp) | Average no. of N nts (bp) |
|---|---|---|---|
| NIPBL$^{+/-}$ | 53.1 | 0.46 | 13.5 |
| SMC1A$^{-}$ | 53.5 | 0.50 | 14.5 |
| Controls | 53.1 | 0.48 | 14.2 |

a. Statistical calculations were performed by Student's t-test. No significant differences were found.
b. The calculations were based on unique sequences.
CDR3: complementary determining region 3; no.: number; bp: base pair; P nts: Palindromic nucleotides; N nts: non-template nucleotides

**Table E3**: Immunological data of CdLS patients P6, P8, P10, P11, SMC1A-1 and SMC1A-2

| | P6 (classical CdLS) | P8 (mild CdLS) | P10 (mild CdLS) | P11 (mild CdLS) | SMC1A-1 | SMC1A-2 |
|---|---|---|---|---|---|---|
| ***Age at sampling*** | **14Y** | **4Y** | **10Y** | **14Y** | **8Y** | **11Y** |
| ***Lymphocytes***[a] | | | | | | |
| Proportion of lymphocytes | 31% (22-50) | NA | 35.5% (20-45) | 33.4% (20-45) | NA | NA |
| Total lymphocytes | 2100 $mm^3$ (1000-5500) | NA | 1700 $mm^3$ (1500-4500) | 1700 $mm^3$ (1500-4500) | 3400 $mm^3$ | 3300 $mm^3$ |
| ***T lymphocytes***[a] | | | | | | |
| T lymphocytes, CD3+ | 84% (52-78) | NA | 66% (63-84) | 70% (65-80) | NA | NA |
| T lymphocytes, CD4+ | 51% (25-48) | NA | 42% (33-57) | **33%↓** (40-50) | NA | NA |
| T lymphocytes, CD8+ | 26% (9-35) | NA | 20% (14-39) | 32% (26-30) | NA | NA |
| CD4/CD8 ratio | 1.96 (0.9-3.4) | NA | 2.1 (1.5-2) | **1.03↓** (1.5-2) | NA | NA |
| T lymphocytes, CD4+ | 1110 $mm^3$ (400-2100) | NA | 715.68 $mm^3$ (400-1535) | 551.1 $mm^3$ (500-800) | NA | NA |
| T lymphocytes, CD8+ | 570 $mm^3$ (200-1200) | NA | 340.8 $mm^3$ (139-1015) | 534.4 $mm^3$ (250-800) | NA | NA |
| ***NK cells, CD56+*** | **5%↓** (6-27) | NA | 8% (7-21) | 5% (5-10) | NA | NA |
| ***B lymphocytes***[a] | | | | | | |
| Proportion B lymphocytes, CD19+ | 8% (8-24) | NA | 19% (7-22) | 19% (10-15) | **12%↓** (13-27) | **5%↓** (13-27) |
| Total B lymphocytes, CD19+ | **160 $mm^3$↓** (200-600) | NA | NA | NA | 408 $mm^3$ (270-860) | **165 $mm^3$↓** (270-860) |
| CD21++CD24+/CD19+ | NA | NA | NA | NA | 84% | 84% |
| CD21+CD24++/CD19+ | NA | NA | NA | NA | 1.7% | 1.2% |
| CD21-CD24-/CD19+ | NA | NA | NA | NA | 5% | 5% |
| CD27+/CD19+ | NA | NA | NA | NA | 31% (>10) | 16% (>10) |
| CD27+IgD+/CD19+ (Marginal zone) | 15% (6-29) | NA | NA | NA | 21% | 8% |
| IgM+IgD+/CD19+ | NA | NA | NA | NA | 78% | 63% |
| IgM+IgD+/CD27+CD19+ | NA | NA | NA | NA | 63% | 37% |
| IgM+IgD-/CD27+CD19+ | NA | NA | NA | NA | 16% | 8% |
| IgM-IgD-/CD27+CD19+ | NA | NA | NA | NA | 17% (>10) | 46% (>10) |
| CD21-CD38-/CD19+ | NA | NA | NA | NA | 5% | 3.7% |
| IgD+CD27- (Naive) | 78% (47-84) | NA | NA | NA | NA | NA |
| IgD-CD27+ (Switched memory) | **4%↓** (8-29) | NA | NA | NA | NA | NA |
| IgM-CD38+ (Plasma blasts) | 1% (0-3.2) | NA | NA | NA | NA | NA |
| IgM++CD38++ (Transitional) | <0.5% (<1) | NA | NA | NA | NA | NA |
| B-CD21-CD38- (CD21 low) | 1% (<4) | NA | NA | NA | NA | NA |
| B cell proliferation[b] | NA | NA | NA | NA | NA | Normal |
| ***Immunoglobulins***[a] | | | | | | |
| IgM | 1.71 g/L (0.27-2.1) | 0.59 g/L (0.27-2.1) | **56.2 mg/dL↓** (60-263) | 65.4 mg/dL (50-268) | NA | 1.47 g/L (0.53-1.62) |
| IgG | 14.5 g/L (6.1-14.5) | 9.65 g/L (6.1-14.5) | **754 mg/dL↓** (768-1632) | 825 mg/dL (768-1632) | NA | 14.97 g/L (6.55-12) |
| IgA | 2.82 g/L (0.7-3.65) | 0.98 g/L (0.5-2.3) | **63.5 mg/dL↓** (68-378) | 180 mg/dL (68-378) | NA | 1.89 g/L (0.5-2.03) |
| Ig switching[c] | NA | NA | NA | NA | NA | **Positive, but weak** |
| ***Specific immunoglobulins*** | | | | | | |
| S-Difteri-ak (IgG) | NA | 0.2 IE/mL (0.01-3.0) | NA | NA | NA | NA |
| S-Tetanus-ak (IgG) | NA | 0.25 IE/mL (0.09-12.87) | NA | NA | NA | NA |
| S-Tetanus-ak (IgG1) | **0.75 mg/L↓** (4.9-180) | 2.9 mg/L (0.9-228.5) | NA | NA | NA | NA |
| S-Haemophilus-ak (IgG) | NA | **0.14 mg/L↓** (0.15-29.5) | NA | NA | NA | NA |
| S-Pneumokock-ak (IgG2) | NA | 11 mg/L (0.8-122.4) | NA | NA | NA | NA |
| S-Pneumokock-ak (IgG) | NA | 25 mg/L (9.2-225.9) | NA | NA | NA | NA |

a. Normal range, from immunological center where measurement was taken, is indicated in parenthesis. Values below the normal range are indicated in bold.
b. Measured after stimulation with IL4, CD40L, CD40L+IL4, NS-BAFF_TH3, BAFF_TH3, BAFF+μ_TH3 or CpG+μ_TH3, NS, non-stimulated; TH3, thymidine incorporation
c. Measured after stimulation with CD40L+IL4 for switching to IgE. NA, not analysed

**Figure E1: V(D)J recombination substrate assays in a human fibroblast cell line and a murine pro-B-cell line.**

a) Schematic picture of the pMX-RSS-GFP/IRES-hCD4 (MX-RSS-INV) construct.

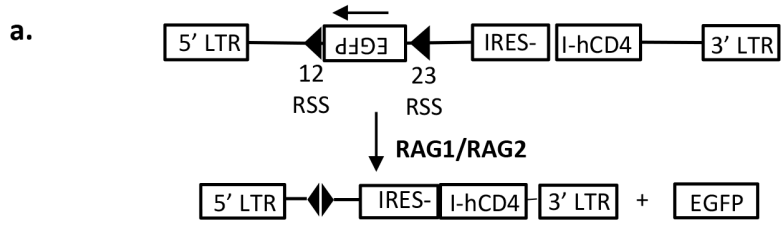
b) Expression of *NIPBL* in a human fibroblast line stably transfected with the MX-RSS-INV) construct and c) expression of *Nipbl* in a mouse pro-B cell line stably transfected with the same construct were measured by real-time PCR. The expression of *NIPBL* or *Nipbl* was reduced by 19% or 36% with the scrambled control siRNAs and by 80% or 74% after specific *NIPBL* or *Nipbl* siRNA treatment.

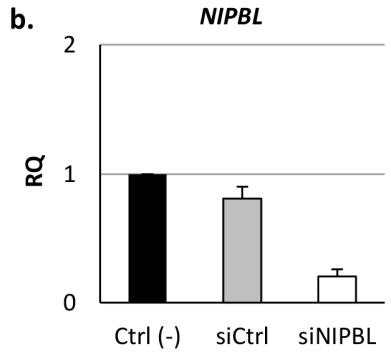
d) Chromosomal-integrated V(D)J reporter assay in human fibroblast cells. The V(D)J reporter MX-RSS-INV was retrovirally transduced into a human fibroblast cell line from an unaffected control. CD4 positive cells were FACS sorted and expanded and tested (see supplemental M&M). GFP positive cells are indicative of successful recombination. Statistical calculation was performed with Student's t-test. The results represent two independent experiments.

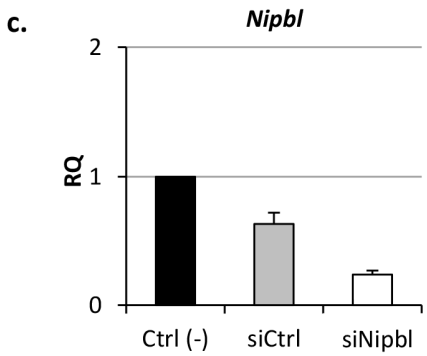
e) V(D)J reporter assay in a murine B-Abl pro-B-cell line carrying the chromosome integrated MX-RSS-INV substrate. GFP positive cells are indicative of successful recombination. Statistical calculation was performed with Student's t-test. The results represent two independent experiments.

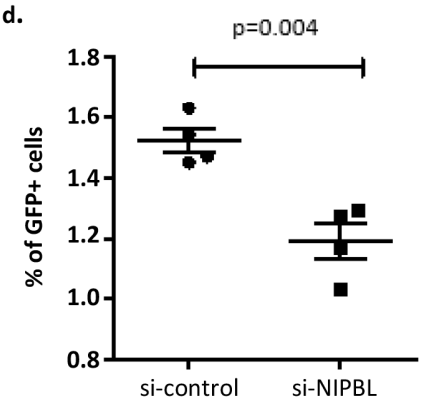
**Fig. E2.** Frequencies of individual *IGHV* genes in unproductive sequences in a) classical *NIPBL* and b) *SMC1A* mutated B-cells from CdLS patients are shown. Statistical calculations were performed by Student's t-test, and significant differences compared to controls are indicated by arrows. ↑ and ↓, increased or decreased compared to controls, respectively.

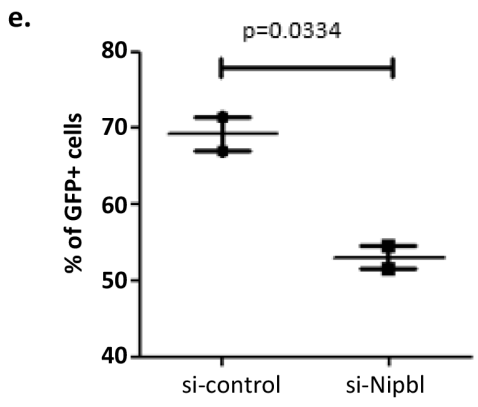
**Fig. E3. Expression of genes analysis in NIPBL knockdown cells.** Expression of *NIPBL* in a human fibroblast line (a) was measured by real-time PCR. The expression of *NIPBL* was reduced by 21% with the scrambled control siRNAs and by 77% after specific *NIPBL* siRNA treatment. b) Expression of genes involved in V(D)J recombination in a human fibroblast cell line. The results represent an average of two independent experiments. RQ: relative quantification; Ctrl: control; (-): untreated


a.
5' LTR
EGFP
IRES-
I-hCD4
3' LTR
12
RSS
23
RSS
RAG1/RAG2
5' LTR
IRES-
I-hCD4
3' LTR
+
EGFP

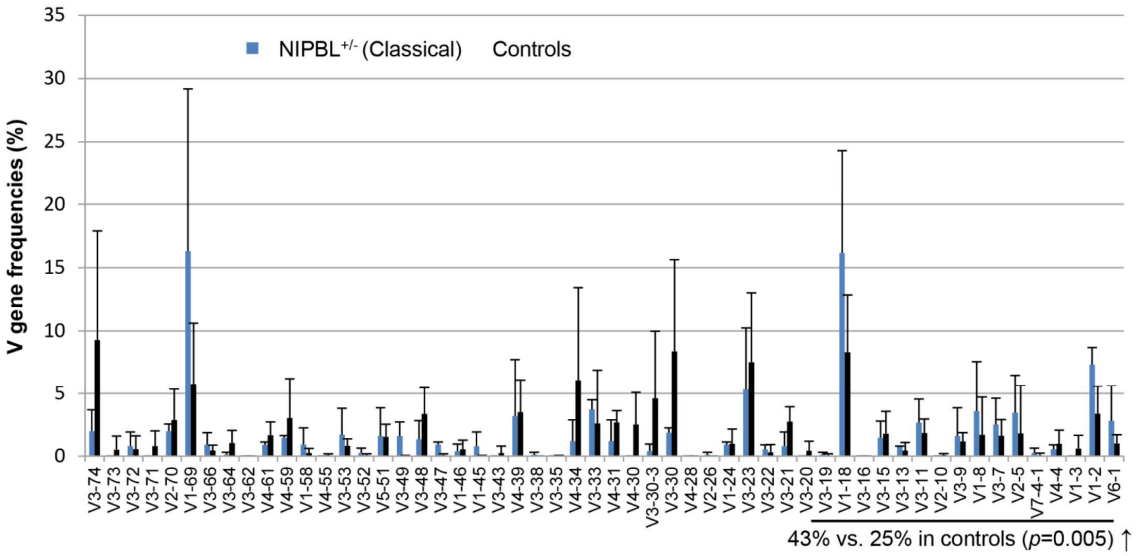

b.
NIPBL
RQ
2
1
0
Ctrl (-)
siCtrl
siNIPBL


c.
Nipbl
RQ
2
1
0
Ctrl (-)
siCtrl
siNipbl

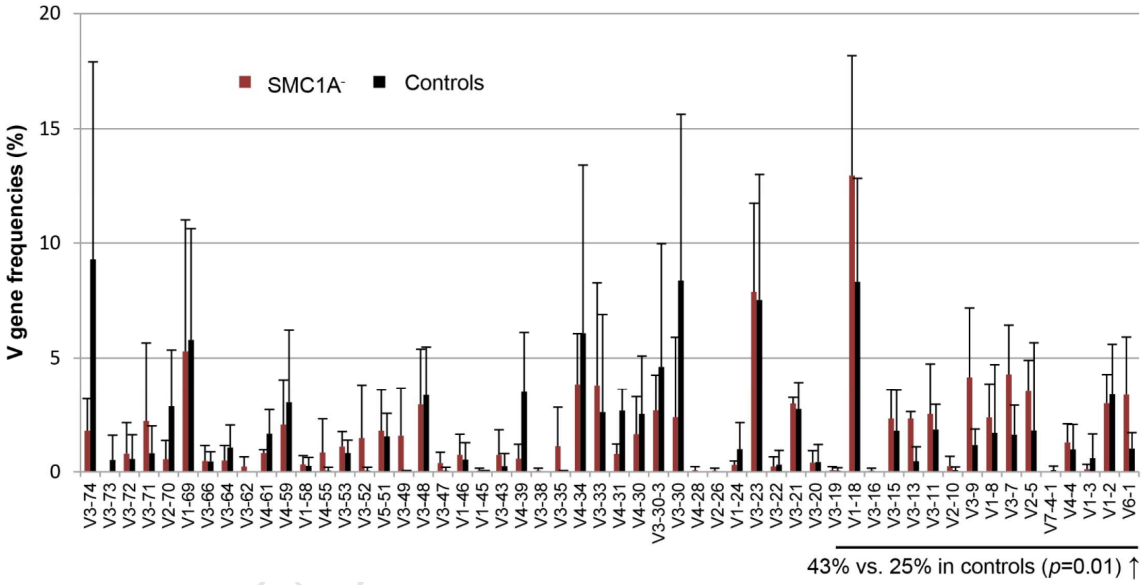

d.
p=0.004
% of GFP+ cells
1.8
1.6
1.4
1.2
1.0
0.8
si-control
si-NIPBL

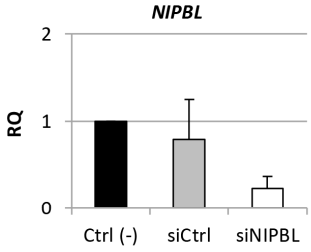

e.
p=0.0334
% of GFP+ cells
80
70
60
50
40
si-control
si-Nipbl

**a.**


NIPBL+/- (Classical)
Controls
V gene frequencies (%)
35
30
25
20
15
10
5
0
V3-74 V3-73 V3-72 V3-71 V2-70 V1-69 V3-66 V3-64 V3-62 V4-61 V4-59 V1-58 V4-55 V3-53 V3-52 V5-51 V3-49 V3-48 V3-47 V1-46 V1-45 V3-43 V4-39 V3-38 V3-35 V4-34 V3-33 V4-31 V4-30 V3-30-3 V3-30 V4-28 V2-26 V1-24 V3-23 V3-22 V3-21 V3-20 V3-19 V1-18 V3-16 V3-15 V3-13 V3-11 V2-10 V3-9 V1-8 V3-7 V2-5 V7-4-1 V4-4 V1-3 V1-2 V6-1
43% vs. 25% in controls (p=0.005) ↑

**b.**


SMC1A-
Controls
V gene frequencies (%)
20
15
10
5
0
V3-74 V3-73 V3-72 V3-71 V2-70 V1-69 V3-66 V3-64 V3-62 V4-61 V4-59 V1-58 V4-55 V3-53 V3-52 V5-51 V3-49 V3-48 V3-47 V1-46 V1-45 V3-43 V4-39 V3-38 V3-35 V4-34 V3-33 V4-31 V4-30 V3-30-3 V3-30 V4-28 V2-26 V1-24 V3-23 V3-22 V3-21 V3-20 V3-19 V1-18 V3-16 V3-15 V3-13 V3-11 V2-10 V3-9 V1-8 V3-7 V2-5 V7-4-1 V4-4 V1-3 V1-2 V6-1
43% vs. 25% in controls (p=0.01) ↑




**V(D)J recombination genes**