Article

New insights into the molecular basis of spinal neurofibromatosis type 1

Paola Bettinaglio[1*], Eleonora Mangano[2*], Viviana Tritto[1], Roberta Bordoni[2], Rosina Paterra[3], Arianna Borghi[3], Marinella Volontè[1], Cristina Battaglia[1,2], Veronica Saletti[4], Claudia Cesaretti[5], Federica Natacci[5], Mariarosa A.B. Melone[6,7], Marica Eoli[3] ° and Paola Riva[1] °

1. Department of Medical Biotechnology and Translational Medicine (BIOMETRA), University of Milan, Segrate, Milan, Italy
2. Institute of Biomedical Technologies (ITB) National Research Center (CNR), ITB-CNR, Segrate, Milan, Italy
3. Molecular Neuroncology Unit, Fondazione IRCCS Istituto Neurologico Carlo Besta, Milan, Italy
4. Developmental Neurology Unit, Fondazione IRCCS Istituto Neurologico Carlo Besta, Milan, Italy
5. Medical Genetics Unit, Woman-Child-Newborn Department, Fondazione IRCCS Ca' Granda-Ospedale Maggiore Policlinico, Milan, Italy
6. Department of Advanced Medical and Surgical Sciences, 2nd Division of Neurology, Center for Rare Diseases and InterUniversity Center for Research in Neurosciences, University of Campania Luigi Vanvitelli, 80131 Naples, Italy
7. Sbarro Institute for Cancer Research and Molecular Medicine, Center for Biotechnology, Temple University, Philadelphia, PA 19122-6078, USA

° Correspondence: Marica Eoli, e-mail: marica.eoli@istituto-besta.it, Telephone number: +39-02-23942285; Paola Riva, e-mail: paola.riva@unimi.it, Telephone number: +39-02-503130462;

*These authors equally contributed to the work

° Co-corresponding authors

Abstract:

Spinal Neurofibromatosis (SNF) is a form of Neurofibromatosis type 1 (NF1) characterized by bilateral neurofibromas involving all spinal roots. The pathogenic mechanisms determining the SNF form are currently unknown. To verify the presence of genetic variants possibly related to SNF or classic NF1, we studied 106 sporadic NF1 and 75 SNF patients using an NGS panel of 286 genes encoding RAS pathway effectors and neurofibromin interactors and evaluated the expression of syndecans (SDC1, SDC2, SDC3, SDC4), the *NF1* 3' tertile interactors, by quantitative real-time PCR. We previously identified 75 and 106 *NF1* variants in SNF and NF1 cohorts, respectively. The analysis of the distribution of pathogenic *NF1* variants in the three *NF1* tertiles showed a significantly higher prevalence of *NF1* 3' tertile mutations in SNF than in the NF1 cohort. We hypothesized a potential pathogenic significance of the 3' tertile *NF1* variants in SNF. The analysis of syndecan expression on PBMCs RNAs from 16 SNF, 16 classic NF1 patients and 16 healthy controls showed that the expression levels of *SDC2* and *SDC3* were higher in SNF and NF1 patients than in controls;

moreover, *SDC2, SDC3* and *SDC4* were significantly over expressed in patients mutated in the 3′ tertile compared to controls. Two different mutational *NF1* spectra seem to characterize SNF and classic NF1, suggesting a pathogenic role of *NF1* 3′ tertile and its interactors, syndecans, in SNF. Our study, providing new insights on a possible role of neurofibromin C-terminal in SNF, could address effective personalized patient management and treatments.

Keywords: SNF; *NF1* 3′ tertile; syndecans

1. Introduction

Neurofibromatosis type 1 (NF1) is a genetic disease caused by usually heterozygous loss-of-function pathogenic variants in the *NF1* gene [1]. The main clinical features are café-au-lait spots, iris Lisch nodules, axillary and inguinal freckles and multiple neurofibromas. Pathogenic variants of *NF1* also cause Spinal Neurofibromatosis (SNF), a peculiar clinical entity belonging to NF1, characterized by bilateral neurofibromas involving all spinal nerve roots, with or without other manifestations of classic NF1. Spinal neurofibromas are found in up to 38% of patients with NF1 and 33% of cases are asymptomatic. In contrast to typical NF1, patients with SNF show a late-onset clinical phenotype, often associated with severe back pain due to multiple symptomatic spinal neurofibromas. SNF patients are characterized by multiple tumor masses symmetrically involving all vertebral levels of the cranial, thoracic and lumbar spine. These features distinguish them from NF1 individuals with spinal neurofibromas in multiple, but not all spinal roots (MNFSR) [2]. SNF can be diagnosed and distinguished from NF1 by spinal MRI; however, the molecular bases of the two forms are still unknown [3].

SNF is also characterized by a high intra familial phenotypic variability: patients affected by it may also belong to families presenting individuals diagnosed with classic NF1 or MNFSR.

63 Up to now, only 98 patients with complete or partial SNF have been described, 90% of them
64 are carriers of a mutation in the *NF1* gene. In a small cohort of SNF patients, it has also been reported
65 that pathogenic missense *NF1* variants are significantly more frequent in the SNF than in the classic
66 patients [2]. The apparent prevalence of missense *NF1* mutations in SNF could be correlated to a
67 gain-of-function significance. According to the complexity of neurofibromin structure reflecting its
68 different functions, it could be interesting to know whether the pathogenic *NF1* variants cluster in
69 specific regions. Sharif et al. [4] proposed the subdivision into tertiles of the *NF1* gene, including
70 specific neurofibromin domains. The middle tertile contains the GAP-Related Domain (GRD) at its
71 N-terminus, that confers to NF1 a RAS pathway inhibitor function specifically related to
72 RASopathies and that includes the tubulin-binding domain (TBD), and the Sec-PH domain at its C-
73 terminus, composed of a Sec14 homologous domain and a pleckstrin homologous domain and
74 involved in phospholipid and membrane protein binding [5]. The 5′ tertile of the gene corresponds
75 to the Cysteine-Serine-Rich Domain (CSRD) at the N-terminal position of the protein,
76 phosphorylated by both Protein Kinase A and Protein Kinase C (PKC); its PKC-dependent
77 phosphorylation increases the RAS-GAP activity of neurofibromin. The 3′ tertile corresponds to the
78 C-Terminal Domain (CTD) that plays an important role in the regulation of the transition from
79 metaphase to anaphase during the cell cycle and contains a nuclear localization signal for the transfer
80 of neurofibromin to the nucleus. Furthermore, this domain interacts with different molecules, such
81 as DPYSL2 (Dihydropyrimidinase-related protein 2), PTK2 (Protein Tyrosine Kinase 2), CASK
82 (Calcium/Calmodulin-dependent Serine protein Kinase) and syndecans, cell surface proteoglycans,
83 for which it is known the neurofibromin binding domain called SBD (Syndecan Binding Domain) or
84 SBR (Syndecan-Binding Region) [6]. Interestingly, the C-terminal end of neurofibromin appears to
85 be the region mainly involved in the self-association into dimers [7].

86 Attempts of correlation between the presence of mutations in the different domains or tertiles
87 of *NF1* and specific clinical manifestations of NF1 have been investigated. Young et al. showed that
88 missense variants of NF1 patients in codons 844 to 848 correlate with a severe phenotype, as these

variants may have a dominant-negative action, destabilizing even wild-type neurofibromin through protein dimerization [8]. Conflicting results have been obtained for the correlation between *NF1* mutations in the 5′ tertile and the risk of optic pathway gliomas (OPGs) development in NF1 patients [4,9]. The complex structure of neurofibromin reflects a complex biological function, mediated by several interactors or modulated by effectors of the RAS pathway and genetic modifiers, that could play a specific role in the onset of the different forms of NF1 as well as in the heterogeneity of the phenotype [6,10–12]. Despite several years of studies with significant efforts to identify modifier genes, the factors that predict disease severity today are little known. Other mechanisms and strategies should be hypothesized to provide clinicians with adequate tools for effective diagnosis, prognosis, and genetic counselling.

We report here a study focused on verifying the presence of the specific mutational spectrum of *NF1*, considering the occurrence of mutation types and their location within the *NF1* gene, in a large cohort of SNF patients [13], including case studies of cohorts described in literature. We also evaluated the occurrence of variants in neurofibromin interactors. Interestingly, potential pathogenic variants of syndecan proteins have been identified mainly in SNF. The results provided could direct future functional studies that could help identify diagnostic markers and pharmacological targets towards personalized medicine.

2. Materials and Methods

2.1. Study subjects and samples collection

The study cohorts consisted of 106 sporadic patients with classic NF1 and 74 patients with SNF of which 56 were sporadic cases and 18 were probands of pure or mixed spinal families, i.e., families in which, in addition to the SNF proband, other relatives with classic form of NF1 or MNFSR were present. Patients were recruited by written informed consents by the IRCSS C. Besta Neurological Institute, by the IRCCS Ca′ Granda Foundation Ospedale Maggiore Policlinico and by Azienda Ospedaliera Universitaria dell′Università degli Studi della Campania “Luigi Vanvitelli”. The

114 protocol of the study was in accordance with the Declaration of Helsinki and was approved by the
115 Fondazione IRCCS Istituto Neurologico Carlo Besta Ethical Committee and Scientific Board (N°50-
116 19/3/2018).

117 Three to five microliters of whole blood were drawn, and the DNAs was extracted from 3 ml of
118 peripheral blood samples using Gentra Puregene Blood Kit (Qiagen).

119 The clinical features of patients and their identification codes of the two cohorts of patients are
120 reported in Paterra. et al [13].

121 *2.2. NGS Analysis*

122 To identify the pathogenic *NF1* variants of the patients' populations, we used two different
123 custom targeted resequencing panels, produced by Agilent Technologies (SureSelect XT panel). The
124 NGStr2 [14] and NGStr3 custom panels include the coding regions (10 bases from the 3' end and 10
125 bases from the 5' end) and the 5' UTR and 3' UTR regions of 285 genes.

126 The list of genes composing the two panels and the protocol for the libraries preparation and
127 Sequencing is described in detail in the file S1 (see Supplementary materials).

128 The Raw reads of NGS data are available in NCBI Short-read Archive (SRA,
129 https://www.ncbi.nlm.nih.gov/sra/) under the accession number PRJNA8509016 and PRJNA688415.

130 *2.3. Variant analysis and interpretation*

131 The functional annotation and impact prediction were performed using ANNOVAR (v.
132 2019Oct24) [15], which includes prediction scores from 20 prediction algorithms and 8 conservation
133 scores from dbNSFP database (https://sites.google.com/site/jpopgen/dbNSFP). ANNOVAR also
134 calculate the Damagepredcount, a value ranging from 0 to 20, indicating how many predictors, out
135 of 20, consider the variant as damaging.

136 Variants with a MAF <0.01 according to both 1000 Genomes database
137 (https://www.internationalgenome.org/, release 20130502) [16] and gnomAD v.3.1.2 were
138 considered rare. In addition, SNVs not reported neither in public databases, such as 1000 Genomes

139 Project, gnomAD v.3.1.2, dbSNP (https://www.ncbi.nlm.nih.gov/snp/, Build 154, April21 2020),
140 DECIPHER v.11.13 (https://decipher.sanger.ac.uk/) [17], and ClinVar
141 (https://www.ncbi.nlm.nih.gov/clinvar) [18], nor in PubMed (https://pubmed.ncbi.nlm.nih.gov)
142 were classified as novel.

143 We then manually assessed the clinical significance of the SNVs according to the American
144 College of Medical Genetics (ACMG)/Association of Molecular Pathology (AMP) guidelines [19],
145 taking into account the novelty of the variant, possible associations of the affected genes with
146 mendelian disorders according to OMIM (Online Mendelian Inheritance in Man) database
147 (https://omim.org) [20], previous inclusion in databases such as DECIPHER v.11.13, ClinVar, and
148 COSMIC (Catalogue of Somatic Mutations in Cancer) v.92 (https://cancer.sanger.ac.uk/cosmic) [21],
149 and/or in PubMed, localization of the variant in functional domains which could be mutational
150 hotspots, *in silico* prediction of pathogenicity based on conservation and type of amino acid
151 substitution.

152 *2.4. Quantitative real- time PCR (qPCR)*

153 The isolation of RNA from the peripheral blood mononuclear cells (PBMCs) of available
154 patients were performed at IRCSS C. Besta Neurological Institute, in the Molecular Neuroncology
155 Laboratory using the "TempusTM Spin RNA Isolation" Kit (Applied Biosystems).

156 The available total RNA (500 ng) was reverse transcribed with Maxima H Minus cDNA
157 Synthesis Master Mix kit with dsDNA (Termofisher). For qPCR assays we selected the *SDC2*, *SDC3*
158 and *SDC4* genes, with an expression level in whole blood greater than 0.5 TPM (transcripts per
159 million, GTEx portal source, https://gtexportal.org). The *SDC1* gene was excluded from the analysis
160 due to its expression level of < 0.5 TPM in the whole blood. Each SYBR Green qPCR assay was
161 performed using GoTaq–qPCR master mix (Promega) and run on a QuantStudio 5 Real-Time PCR
162 Systems (Thermo Fisher Scientific). An accurate design using Primer3 (https://primer3.ut.ee) of the
163 oligonucleotides were performed (table S2) to amplify and detect only the target sequences of
164 interest. Three pairs of oligonucleotides were obtained, in each pair one of the oligonucleotides was

designed between an exon-exon junction to reduce the risk of amplifying gDNA in a non-specific way.

2.5. Statistical analysis

The X^2 test or Fisher exact test was used to compare categorical variables. The Benjamini-Hochberg (B-H) method with false discovery rates of 0.05, 0.025 and 0.01 was used to correct p-values for multiple testing. A p value <0.05 was considered as statistically significant. The X^2 test or Fisher exact test was performed using the tools available at https://www.socscistatistics.com/tests. The B-H correction was performed using the tool available at https://tools.carbocation.com/FDR. The distribution of variants in patients with classic NF1 and SNF into three *NF1* tertiles was established accordingly to the subdivision of the *NF1* gene reported in the literature [9]. Specifically, the *NF1* mutations located into the exons 1-21 (amino acids 1-950, including the CSRD) were assigned to the 5′ tertile, that located into the exons 22-38 (amino acids 951-1916, involving the GRD and Sec-PH domains) to the middle tertile, and that into the exons 39 to 57 (amino acids 1917-2818, including the CTD) to the 3′ tertile (transcript ID: ENST00000356175.3 at https://www.ensembl.org; protein ID: NP_000258.1 at https://www.ncbi.nlm.nih.gov/protein).

All qPCR experiments were run in triplicate and the average of the threshold cycles (Ct) for each sample was made. To determine the relative gene expression, the $2^{-\Delta Ct}$ method was applied (ΔCt= Ct gene target – Ct housekeeping gene, for each sample). For each gene analyzed, mean, standard deviation, standard error of the mean, and confidence intervals values were calculated in the three groups of samples, which include 16 patients with SNF, 16 patients with classic NF1 and 16 healthy controls. The equal variance Student's *t*-test was applied to compare the means and the B-H correction for multiple tests with false discovery rate of 0.05, 0.025 and 0.01 was applied. The outliers' values, identified by Tukey test with k= 1.5, were excluded from the analysis. The results were considered statistically significant when $p < 0.05$. The Statistical analysis was performed using the tools available at https://www.socscistatistics.com/tests.

190 *2.6. NF1 interactors selection*

191 The NF1 interactors were selected by means of the IntAct tool (https://www.ebi.ac.uk/intact)
192 between the interactors with an experimentally proven interaction with NF1 obtained by
193 socioaffinity inference, two hybrid, anti-tag coip, anti-bait coip, crosslink and two hybrid pooling
194 method. They were also selected because of the evidence collected in the review by Ratner et al. [6].

195 **3. Results**

196 *3.1. Prevalence of NF1 3′ tertile mutations in SNF vs classic NF1*

197 With the aim of verifying a possible prevalent localization of the mutations along the *NF1* gene,
198 we subdivided the *NF1* gene into tertiles, and evaluated the occurrence of the mutations in SNF and
199 in classic NF1.

200 The prevalence of mutations occurring in the 3′ tertile of the *NF1* gene is significantly higher in
201 patients with SNF (34.4%) than that observed in classic patients (16%) (p=0.006; OR 2.277; CI = 1.31–
202 5.7), while the prevalence of those in the middle tertile was lower (p=0.038; OR 0.49; CI = 0.25–0.96)
203 (Table 1). The localization of the *NF1* 3′ tertile mutations within the specific domains of the C-
204 terminal part of the neurofibromin are reported in Table S3.

205 Since SNF is a rare form of NF1 and relatively few patients with SNF have been described, we
206 performed a combined analysis aimed at verifying the occurrence of specific localization of *NF1*
207 mutations by pooling data obtained in our SNF cohort with those of SNF patients reported in the
208 literature. Considering the 49 selected SNF patients described by Ruggieri [2] after applying
209 stringent diagnostic criteria and knowing that relatives are also included in this cohort, we counted
210 the variants shown by unrelated SNF patients, reducing the number to 25. After merging the
211 mutational data of the 25 reported unrelated SNF patients with those of our 75 unrelated SNF
212 patients, we determined the occurrence of each pathogenic variant from the *NF1* tertiles and
213 compared them to those occurring in 106 classic NF1 patients. Pooled analysis of the distribution of
214 *NF1* variants in the tertiles confirmed a significantly higher prevalence of pathogenic *NF1* variants

in the 3′ tertile in SNF compared to classic NF1, patients with an increase in statistical significance (P=0.0016), with a false discovery rate of 0.05, 0.025 and 0.01 after correction for multiple testing using Benjamin Hochberg's procedure.

3.2. Distribution of different classes of variants within the NF1 tertiles

The prevalence in the 5′ tertile of frameshift and pathogenic missense variants differs significantly, even after B-H correction (p=0.00619 and p=0.0045 respectively, table S4), between classic NF1 and SNF patients (Figure 1a). The distribution in the middle tertile showed no statistical differences between the two groups of patients (Figure 1b). The prevalence of pathogenic 3′ tertile *NF1* stop-gain variants was lower in SNF (Figure 1c) than in NF1 classic patients (p= 0.0127) and remains significant after B-H correction. In the 3′ tertile we found no pathogenic missense variants in classic (0%) versus SNF (5%) patients. For this reason, we could not apply the Chi-square or Fisher's exact test. .

3.3 Neurofibromin interactor variants in SNF and classic NF1 patients

We hypothesized a functional significance of the prevalence of pathogenic *NF1* 3′ tertile variants in SNF. Accordingly, we verified the presence of variants in syndecans, which bind the neurofibromin SBR, in SNF and NF1 patients. These interactors are encoded by genes belonging to the syndecan family: *SDC1, SDC2, SDC3* and *SDC4*. We searched for rare variants with MAF <0.01 by ANNOVAR annotation (Table S5). Six variants in the four genes encoding syndecans were identified in five SNF patients and in one classic NF1 patient (Table 2). We evaluate the clinical significance according to the ACMG/AMP criteria, which led to their classification into two groups (Table S6): 1 "Uncertain" (4/6, 67%), when the evidence was not sufficient to draw definitive conclusions on pathogenicity, including C215T (p.(T72M)) in *SDC1,* C923T (p. (P308L)) and c. A721G (p.(T241A)) in *SDC3*, c. A92G (p. (D31G)) in *SDC4*. 4. "Probably pathogenic" (2/6, 33%), when the evidence supporting pathogenicity was concordant across several different *in silico* predictors although at least one major pathogenicity criterion, such as detection in other patients with similar

240 phenotypes or functional validation of the variant, was still missing. This group includes variants
241 G830A (p.(R277H)) and c. T449C (p.(I150T)) affecting the *SDC1* and *SDC2* genes, respectively. G830A
242 (p.(R277H)) in *SDC1* has never been reported in gnomAD v.3.1.1 and 1000 Genomes database and
243 both variants are found in the cytoplasmic syndecan domain and are predicted to be harmful by
244 most predictors (18 out of 20). All but one of the syndecan variants were co-present in patients
245 carrying pathogenic variants of the 5′ or middle *NF1* tertile. Two of the six variants were predicted
246 to be harmful by most of the predictors questioned by Annovar (Dampred= 18.2) (Table S6).
247 We also searched for variants in *NF1* 5′ and middle tertile interactors, applying the pipeline
248 described above and found 4 rare variants. Applying the criteria described above, we classified the
249 4 variants into two groups (Table S7): 1 "Uncertain", including one variant in *SPRED1*, a GRD
250 domain interactor, found in one classic patient. 2. "Probably pathogenic" including three variants in
251 the *APP* gene, a GRD domain interactor, in two classic patients (1214 and 1165) and in one SNF
252 patient (1085). Interestingly, the SNF patient does not carry pathogenic *NF1* variants.

253 *3.4. The expression of syndecan transcripts in SNF and classic NF1*

254 We studied the expression of syndecans genes in SNF and NF1 patients by qPCR on PBMCs'
255 RNA samples from 16 SNF patients, 16 classic NF1 patients and 16 healthy controls. For qPCR assays
256 we selected the genes of *SDC2, SDC3* and *SDC4* with an expression level in whole blood higher than
257 0.5 TPM (Transcripts Per Million). The *SDC1* gene was not expressed in SNF and NF1 patients as
258 expected, according to data reported by GTex for whole blood, being <0.5 TPM. The average value
259 of the quantitative expression levels ($2^{-\Delta Ct}$) of *SDC2* and *SDC3* was significantly higher in SNF and
260 NF1 patients than in controls, after application of Student's *t*-test and the B-H correction for multiple
261 tests (Table S8 and Figure 2 a and b). The average value of the quantitative expression levels ($2^{-\Delta Ct}$)
262 of *SDC4* was significantly higher in classic NF1 patients than in controls (Table S8 and Figure 2 c).
263 Since both SNF patients and classic NF1 ones showed increased levels of syndecans' transcripts
264 compared with controls, we investigate whether this overexpression was possibly associated with
265 the presence of pathogenic *NF1* variants in a specific tertile of *NF1*, without distinguishing between

266 the specific forms of NF1. We compared the average value of the quantitative expression levels ($2^{-\Delta Ct}$)
267 of the *SDC2*, *SDC3* and *SDC4* in 39 NF1 patients (we added 7 NF1 patients to the 32 previously
268 analyzed to enlarge the cohort) divided into three subgroups according to the specific *NF1* tertile.
269 *SDC2*, *SDC3* and *SDC4* were significantly overexpressed in patients with pathogenic *NF1*
270 variants in the 3′ tertile compared with controls (Figure 3). Furthermore, the *SDC2* and *SDC4* genes
271 were significantly overexpressed in patients with pathogenic *NF1* variants in the 3′ tertile compared
272 with patients with pathogenic *NF1* variants in the middle tertile (Table S9 and Figure 3 a and c).
273 These data suggest that the presence of pathogenic *NF1* variants in the 3′ tertile, including the SBR
274 domain, is probably associated with the increase of *SDC2* and *SDC4* expression.

275 **4. Discussion**

276 SNF is a distinct clinical entity of NF1. It can be distinguished from classic NF1 by spinal MRI,
277 however the molecular bases of the two forms are still unknown [3]. Recently, previous observations
278 indicating the prevalence of pathogenic *NF1* missense variants in SNF were confirmed in a large
279 SNF cohort [2,13]. The results provided by NGS targeted resequencing in a large cohort, reported by
280 Paterra et al. [13], indicate that there is a statistically significant difference in the frequency of
281 pathogenic missense variants between classic and spinal patients, as previously reported [2]. With
282 the present work we increased the significance of the statistical analysis by merging our SNF cohort
283 and the SNF patients clinically and genetically described in the literature [2], by means of a combined
284 analysis. Our data indicate that SNF and classic NF1 are characterized by two different mutational
285 spectra, specifically enriched in pathogenic missense variants in SNF patients. The presence of
286 pathogenic *NF1* missense variants could indicate that their functional significance could lead to a
287 dominant-negative action that destabilizes also the wild-type neurofibromin through protein
288 dimerization or a gain-of-function of mutant neurofibromin that impairs additional pathways other
289 than RAS signaling activation, typically involving loss-of-function of the pathogenic *NF1* variants
290 underlying classic NF1. The truncating and frameshift pathogenic variants, proportionally more
291 frequent in classic patients, lead to a loss of protein function, while the missense pathogenic variants,

292 observed more frequently in patients with spinal form, could lead to an increase in neurofibromin
293 functions.

294 The demonstration that full-length neurofibromin dimerizes with high affinity both *in vitro* and
295 in human cells [7] and the characterization of the protein regions specifically involved in dimer
296 formation, through cryo-electron microscope (Cryo-EM) studies carried out on isoform I [22] and
297 isoform II [23] of neurofibromin, have provided a further insight into the possible role of *NF1*
298 variants. The CryoEM structures of neurofibromin showed that the C-HEAT α-helical domains at
299 the 3′ tertile of both the dimer chains, consisting of HEAT-like repeats commonly involved in
300 protein/protein interactions, form the primary dimer interface. Furthermore, two smaller dimer
301 interfaces symmetrically link the N-terminal of one chain with the C-terminal of the second chain,
302 at both the protein ends. Since it has been demonstrated in several studies that heterozygous cells
303 with truncated *NF1* or frameshift mutations show an amount of neurofibromin less than 50%,
304 Sherekar et al. hypothesized that the aforementioned types of mutation are involved in the
305 degradation of the wild-type protein, which forms a dimer with the mutated protein [7]. This
306 mechanism may not always be applied to pathogenic missense variants found in spinal patients.
307 Consistently, in 2019 Frayling et al. [24] and in 2023 Young et al. [8] hypothesized that *NF1* missense
308 mutation could affect the function of neurofibromin which acts within the cell as a dimer, probably
309 acting in a dominant negative manner.

310 The data presented here led us to formulate two hypotheses: the impairment of *NF1* 3′ tertile
311 may i) be less associated with classic NF1 or ii) have a pathogenic significance in SNF. The prevalence
312 of pathogenic variants in the 3′ tertile, including the majority of the dimerization domain, could lead
313 to increased or decreased efficiency in dimer formation, with a possible gain or loss of protein
314 functions [7]. The coding regions of the 3′ tertile, where the pathogenic variants of spinal cases occur
315 more frequently than the classic ones, correspond to the C-terminal of neurofibromin. HLR (HEAT-
316 like Repeat Region), NLS (Nuclear Localisation Signal) and SBR are present in this protein portion.
317 The first two domains are necessary for nuclear localization of neurofibromin, while SBR is involved

318 in the translocation of neurofibromin along the membrane, binding to syndecans. The interaction
319 between neurofibromin and syndecans is important for cell differentiation and proliferation and for
320 synaptic plasticity [25]. Interestingly, three out of 39 (8%) *NF1* mutations of the 3′ tertile were in the
321 SBR domain, 31 out of 39 (79%) in the HLR region, and 4 out of 39 (10%) in regions of the CTD
322 domain not belonging to HLR, NLS or SBR (Table S3). As for the splicing mutation *NF1*: c.8051-1G>C
323 it was not possible to predict the localization. Most of the 3′ *NF1* tertile mutations (27 out of 39, 70%)
324 were truncating and located in the HLR (Table S3), thus causing the partial or complete loss of the
325 HLR, or the whole loss of NLS and SBR domains of neurofibromin. The clusterization of *NF1*
326 truncating mutations in the CTD of the protein could have a structural effect on the dimer formation
327 and a functional effect on the binding of NF1 to the syndecans or to other interactors of the CTD
328 domain [7,23]. We hypothesize that in addition to pathogenic variants in the 3′ tertile also interactor
329 variants of the included neurofibromin domains might play a role in the development of SNF rather
330 than classic NF1.

331 Colocalization of neurofibromin near the pre- or post-synaptic membrane could promote the
332 GTPase activity of RAS or contribute to specific NF1 phenotypes, currently unknown. Single cell
333 RNAseq data show that the levels of distribution in different cell types of *NF1* and syndecan
334 transcripts are comparable (Gtex portal). Furthermore, the four syndecans in turn interact with
335 CASK, a membrane-associated guanylate cyclase expressed in the embryonic and postnatal brain.
336 The binding with neurofibromin causes the formation of the neurofibromin-syndecan-CASK, a
337 ternary protein complex, whose role has not yet been identified [25]. Given this evidence from the
338 literature, we investigated the presence of genetic variants in the *NF1* 3′ tertile interactors and found
339 the coexistence of syndecan genetic variants with an uncertain and probable clinical pathogenic
340 significance only in SNF patients carrying pathogenic variants in the other two *NF1* tertiles. No
341 classic patient presents this type of variants. We currently do not know whether the presence of
342 missense variants in the 3′ tertile of the *NF1* gene or the coexistence of variants in syndecans with
343 pathogenic *NF1* variants in the 5′ and middle tertile could help determine the spinal cord phenotype

344 and elucidate the role of syndecans in NF1, but the possible role of this *NF1* region may be a
345 challenge for studies aimed at identifying not only new diagnostic markers, but also
346 pharmacological targets in NF1 disease. Considering the distribution of pathogenic variants in the
347 *NF1* tertiles in the SNF compared to classic NF1 and the syndecan variants detected in the SNF, we
348 lean towards the second hypothesis, namely that the prevalence of *NF1* missense mutations and the
349 preferential localization of *NF1* mutations in the 3′ tertile in SNF rather than in classic NF1 may
350 contribute to the pathogenesis of SNF. Finally, we found that syndecans are overexpressed in SNF
351 and classic NF1, suggesting their possible role in NF1 pathogenesis. The finding that syndecans are
352 overexpressed when neurofibromin function is impaired in both SNF patients and classic NF1 ones
353 and that the highest values in syndecan expression were found in patients with pathogenic *NF1*
354 variants in the 3′ tertile consistent with their involvement in the activity and correct functionality of
355 the complex formed with neurofibromin. Knowing that syndecans are adhesion molecules, their
356 functions could be inhibited by specific pharmacological treatments. Interestingly, monoclonal
357 antibodies and syndecan enzyme inhibitors are extensively studied in various types of cancer with
358 inhibiting effects on cell growth and migration [26]. These molecules could also be used in induced
359 pluripotent stem cells derived from NF1 patients to reduce the high level of syndecans observed in
360 patients, to evaluate the role of syndecans in disease and to identify potential pharmacological
361 treatments for NF1.

362 Two different *NF1* mutational spectra seem to characterize SNF and classic NF1, suggesting the
363 implication of the *NF1* 3′ tertile NF1 and syndecans in SNF. There are no data in the literature that
364 confirm or help us understand the syndecan deregulation we observed in NF1. The data provided
365 here should direct functional studies and understand the mechanisms underlying the deregulation
366 of syndecans in NF1 and SNF and the implication of specific neurofibromin domains as well as its
367 interactors. This is a pilot study that, by providing new insights into the complex role of
368 neurofibromin in NF1, could address future investigations aimed at identifying the pathogenic

causes of specific forms of NF1, promoting the development of a personalized medicine in both diagnostic and pharmacological fields.

Supplementary Materials: File S1: NGS method and panels, Table S2: Primers used for the Real-Time PCR reaction, Table S3: Localization of the *NF1* 3′ tertile mutations within the neurofibromin domains, Table S4: Statistical analysis of variants distribution in tertiles, Table S5: ANNOVAR annotation of rare syndecans variants, Table S6: Evaluation of the clinical significance of rare syndecan variants, Table S7: Evaluation of the clinical significance of interactors rare variants, Table S8: Statistical analysis of Real time PCR data, Table S9: Statistical analysis of Real time PCR data by *NF1* tertiles.

Author Contributions: Conceptualization, P.R., M.E., P.B.; methodology, P.B., V.T., R.B., R.P., A.B., E.M.; software, E.M.; validation, V.T., P.B., E.M., E.M., C.C., M.A.B.M.; investigation, P.B., E.M., P.R.; resources M.E., P.R., F.N.; data curation, F.N., P.B., M.E.; writing—original draft preparation, P.R., P.B., E.M.; writing—review and editing, P.R., P.B.; supervision, P.R., M.E., F.N.; project administration, P.R.; funding acquisition, M.E., P.R., F.N, M.A.B.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Italian Ministry of Health, grant number RF-2016-02361293 (to P.R., M.E., F.N.), by ANF ODV (to P.R.), and by University of Campania "Luigi Vanvitelli" - project V:ALERE 2019 Id343-TRANSITION "Nutri-epigenetics and physical activity: a natural help for Neurofibromatosis type 1" (to M.A.B.M.).

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Fondazione IRCCS Istituto Neurologico Carlo Besta Ethical Committee and Scientific Board (N°50- 19/3/2018).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The raw reads of panel NGStr3 are available in NCBI Short-read Archive (SRA, https://www.ncbi.nlm.nih.gov/sra) under the accession number PRJNA8509016 and the raw reads of panel NGStr2 under the accession number PRJNA688415.

Acknowledgments: The authors are grateful to patients for their cooperation and support. The authors thank Simona Curti for English language revision. This research is supported (not financially) by the European Reference Network on Genetic Tumour Risk Syndromes (ERN GENTURIS). The authors M.E, V.S, C.C and F.N are members of ERN GENTURIS. ERN GENTURIS is funded by the European Union.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kang E, Kim YM, Seo GH, Oh A, Yoon HM, Ra YS, et al. Phenotype categorization of neurofibromatosis type I and correlation to NF1 mutation types. *J. Hum. Genet.* 2020;65:79–89.

2. Ruggieri M, Polizzi A, Spalice A, Salpietro V, Caltabiano R, D'Orazi V, et al. The natural history of spinal neurofibromatosis: a critical review of clinical and genetic features. *Clin. Genet.* 2015;87:401–410.

3. Nguyen R, Dombi E, Akshintala S, Baldwin A, Widemann BC. Characterization of spinal findings in children and adults with neurofibromatosis type 1 enrolled in a natural history study using magnetic resonance imaging. *J. Neurooncol.* 2015;121:209–215.

4. Sharif S, Upadhyaya M, Ferner R, Majounie E, Shenton A, Baser M, et al. A molecular analysis of individuals

411 with neurofibromatosis type 1 (NF1) and optic pathway gliomas (OPGs), and an assessment of genotype-
412 phenotype correlations. *J. Med. Genet.* 2011;48:256–260.

413 5. Bergoug M, Doudeau M, Godin F, Mosrin C, Vallée B, Bénédetti H. Neurofibromin Structure, Functions and
414 Regulation. *Cells*. 2020;9.

415 6. Ratner N, Miller SJ. A RASopathy gene commonly mutated in cancer: the neurofibromatosis type 1 tumour
416 suppressor. *Nat. Rev. Cancer*. 2015;15:290–301.

417 7. Sherekar M, Han S-W, Ghirlando R, Drew M, Rabara D, Waybright T, et al. Biochemical and structural analyses
418 reveal that the tumor suppressor neurofibromin (NF1) forms a high-affinity dimer. *J. Biol. Chem.* 2020;295:1105–
419 1119.

420 8. Young LC, Goldstein de Salazar R, Han S-W, Huang ZYS, Merk A, Drew M, et al. Destabilizing NF1 variants act
421 in a dominant negative manner through neurofibromin dimerization. *Proc. Natl. Acad. Sci. U. S. A.*
422 2023;120:e2208960120.

423 9. Melloni G, Eoli M, Cesaretti C, Bianchessi D, Ibba MC, Esposito S, et al. Risk of Optic Pathway Glioma in
424 Neurofibromatosis Type 1: No Evidence of Genotype-Phenotype Correlations in A Large Independent Cohort.
425 *Cancers (Basel).* 2019;11.

426 10. Wang W, Wei C-J, Cui X-W, Li YH, Gu YH, Gu B, et al. Impacts of NF1 Gene Mutations and Genetic Modifiers
427 in Neurofibromatosis Type 1. *Front. Neurol.* 2021;12:704639.

428 11. Mußotter T, Kluwe L, Högel J, Nguyen R, Cooper DN, Mautner VF, et al. Non-coding RNA ANRIL and the
429 number of plexiform neurofibromas in patients with NF1 microdeletions. *BMC Med. Genet.* 2012;13:98.

430 12. Tritto V, Ferrari L, Esposito S, Zuccotti P, Bianchessi D, Natacci F, et al. Non-Coding RNA and Tumor
431 Development in Neurofibromatosis Type 1: ANRIL Rs2151280 Is Associated with Optic Glioma Development
432 and a Mild Phenotype in Neurofibromatosis Type 1 Patients. *Genes (Basel).* 2019;10.

433 13. Paterra R, Bettinaglio P, Borghi A, Mangano E, Tritto V, Cesaretti C, et al. A Translational Approach to Spinal
434 Neurofibromatosis: Clinical and Molecular Insights from a Wide Italian Cohort. *Cancers (Basel).* 2022;15.

435 14. Ferrari L, Mangano E, Bonati MT, Monterosso I, Capitanio D, Chiappori F, et al. Digenic inheritance of subclinical
436 variants in Noonan Syndrome patients: an alternative pathogenic model? *Eur. J. Hum. Genet.* 2020;28:1432–1445.

437 15. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput
438 sequencing data. *Nucleic Acids Res.* 2010;38:e164.

439 16. Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, et al. A global reference for human genetic
440 variation. *Nature*. 2015;526:68–74.

441 17. Firth HV, Richards SM, Bevan AP, Clayton S, Corpas M, Rajan D, et al. DECIPHER: Database of Chromosomal
442 Imbalance and Phenotype in Humans Using Ensembl Resources. *Am. J. Hum. Genet.* 2009;84:524–533.

443 18. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, et al. ClinVar: improving access to variant
444 interpretations and supporting evidence. *Nucleic Acids Res.* 2018;46:D1062–D1067.

445 19. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation
446 of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and
447 Genomics and the Association for Molecular Pathology. *Genet. Med. Off. J. Am. Coll. Med. Genet.* 2015;17:405–
448 424.

449 20. Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA. Online Mendelian Inheritance in Man (OMIM),
450 a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.* 2005;33:D514-7.

451 21. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: the Catalogue Of Somatic
452 Mutations In Cancer. *Nucleic Acids Res.* 2019;47:D941–D947.

453 22. Lupton CJ, Bayly-Jones C, D'Andrea L, Huang C, Schittenhelm RB, Venugopal H, et al. The cryo-EM structure
454 of the human neurofibromin dimer reveals the molecular basis for neurofibromatosis type 1. *Nat. Struct. Mol.*
455 *Biol.* 2021;28:982–988.

456 23. Naschberger A, Baradaran R, Rupp B, Carroni M. The structure of neurofibromin isoform 2 reveals different
457 functional states. *Nature*. 2021;599:315–319.

458 24. Frayling IM, Mautner V-F, van Minkelen R, Kallionpaa RA, Aktaş S, Baralle D, et al. Breast cancer risk in
459 neurofibromatosis type 1 is a function of the type of NF1 gene mutation: a new genotype-phenotype correlation.
460 *J. Med. Genet.* 2019;56:209–219.

461 25. Hu H-T, Shih P-Y, Shih Y-T, Hsueh Y-P. The Involvement of Neuron-Specific Factors in Dendritic Spinogenesis:
462 Molecular Regulation and Association with Neurological Disorders. *Neural Plast.* 2016;2016:5136286.

463 26. Malavaki CJ, Roussidis AE, Gialeli C, Kletsas D, Tsegenidis T, Theocharis AD, et al. Imatinib as a key inhibitor
464 of the platelet-derived growth factor receptor mediated expression of cell surface heparan sulfate proteoglycans
465 and functional properties of breast cancer cells. *FEBS J.* 2013;280:2477–2489.

466

467

468

469 **Legend to figures**

470 **Figure 1.** Distribution of splicing, stopgain, frameshift insertion-deletion, non-frameshift insertion-
471 deletion, missense, large deletion (LD) and of *NF1* gene mutations in the 5′ (a), middle (b) and 3′ (c)
472 tertile of the *NF1* gene. Statistically significant P-values obtained by Fisher's exact test or Chi square and
473 after correction for multiple tests using Benjamin Hochberg procedure are shown above the bars.

474 # significant with an FDR of 0,05 and 0,025 after B-H correction for multiple tests

475 * significant with an FDR of 0,05 after B-H correction for multiple tests

476

477 **Figure 2.** Box plots show the dispersion and quantitative expression levels of gene expression values
478 ($2^{-\Delta Ct}$) analyzed by qPCR of the syndecans genes *SDC2, SDC3* and *SDC4* in PBMCs of 16 patients with
479 SNF (shown in black), 16 controls (WT, shown in white) and 16 patients with classic NF1 (shown in
480 grey). *SDC2, SDC3* and *SDC4* were statistically significantly overexpressed, even after B-H correction
481 for multiple tests, in SNF and classic NF1 patients compared with controls, Student's t-test. The boxes
482 represent the 25th and 75th percentiles. The whiskers show the minimum and maximum value of the

483 distribution, excluding outliers. The broad horizontal lines represent the median value. Outliers are
484 represented as points outside the boxes and excluded from the Student's t-test analysis. Statistically
485 significant P.values obtained from Student's t-test and after correction for multiple tests using Benjamin
486 Hochberg procedure are shown above the bars.

487 # significant with an FDR of 0.05 and 0.025 after B-H correction for multiple tests

488 * significant with an FDR of 0.05 after B-H correction for multiple tests

489 § significant with an FDR of 0.05, 0.025 and 0.01 after B-H correction for multiple tests

490

491 **Figure 3.** Syndecan expression in different *NF1* mutated tertiles in patients and controls. The box plots
492 show the dispersion and the quantitative expression levels of the gene expression values ($2^{-\Delta Ct}$) analyzed
493 by qPCR of the syndecans genes *SDC2* (a), *SDC3* (b) and *SDC4* (c) in PBMCs from 18 patients with *NF1*
494 mutations in the 5′ tertile (5′, shown in black), 11 patients with *NF1* mutations in the middle tertile
495 (middle, shown in light grey), 10 patients with *NF1* mutations in the 3′ tertile (3′, shown in grey) and 16
496 healthy controls (WT, shown in white). *SDC2*, *SDC3* and *SDC4* were statistically significantly hyper-
497 expressed, even after B-H correction for multiple tests, in patients with *NF1* mutations of the 3′ tertile as
498 compared to controls. *SDC2* and *SDC4* were statistically significantly hyper-expressed in patients with
499 *NF1* mutations of the 3′ tertile as compared with patients carrying *NF1* mutations in the middle tertile,
500 Student's t-test. The boxes represent the 25th and 75th percentiles. The whiskers show the minimum and
501 maximum value of the distribution, excluding the outliers. The big horizontal lines represent the median
502 value. The outliers are represented as spots outside of the boxes and excluded from the Student's t-test
503 analysis. Statistically significant P values obtained by Student's t-test and after correction for multiple
504 testing using Benjamin Hochberg procedure are showed above the bars.

505 # significant with a FDR of 0.05 and 0.025 after B-H correction for multiple tests

506 § significant with a FDR of 0.05, 0.025 and 0.01 after B-H correction for multiple tests

507