Supplemental Data

AJHG*, Volume *86

Targeted Next-Generation Sequencing Appoints *C16orf57* as Clericuzio-Type Poikiloderma with Neutropenia Gene

Ludovica Volpi, Gaia Roversi, Elisa Adele Colombo, Nico Leijsten, Daniela Concolino, Andrea Calabria, Maria Antonietta Mencarelli, Michele Fimiani, Fabio Macciardi, Rolph Pfundt, Eric F.P.M. Schoenmakers, and Lidia Larizza

Figure S1. Flow Chart of the Array Capture and Deep Sequencing Procedure

5 μg high molecular weight DNA from patient V-2

Fasteris SA (CH)

- DNA fragmentation, ligation of paired-end adapters and selection of 200 bp fragments
- PCR amplification (14 cycles) to generate >20 μg template library to be captured, according to Illumina guidelines
- Quality control of the DNA Colony Template Library by cloning an aliquot into TOPO Plasmid and capillary sequencing of 5-10 clones

Agilent

- Custom Repeats-masked 244K Array chip (3.4 Mb region cut down to 1.7 Mb)
- 4 bp tiling between neighboring oligos in repeat-free regions

ImaGenes GmbH (Germany)

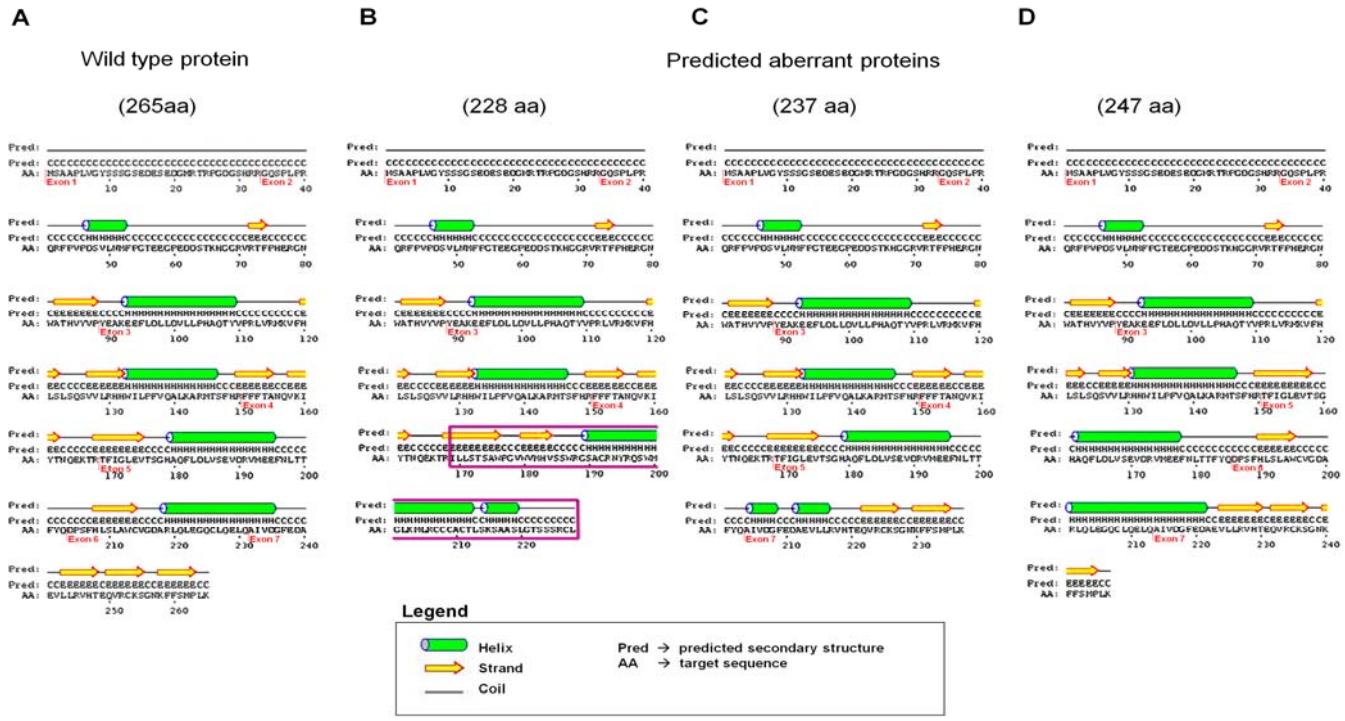
- Hybridization cocktail according to Agilent's CGH protocol for 244K array without Cot DNA
- Incubation: 3 min @ 95°C, 30 min @37°C
- Hybridization to a 244K Agilent array 72h @65°C, 20 rpm
- Washes according to Agilent's CGH protocol:
  Wash buffer I: 5 min @ RT
  Wash buffer II: 1 min @ 37°C
- Elution of captured DNA
  3 x 5 min @ 95°C with water
- DNA concentration by speed vacuum

Fasteris SA (CH)

- Captured DNA amplified for 10 cycles
- Quality control of the DNA Colony Template Library by cloning an aliquot into TOPO Plasmid and capillary sequencing of 5-10 clones
- High-throughput DNA sequencing on Illumina Genome Analyzer GAII for 36 cycles (Solexa technology)
- Bioinformatics analysis
  ELAND mapping V.GAPipeline-1.0 vs Homo sapiens
  MAQ II (V.0.7.1) vs target region
- Enrichment = $\frac{\Sigma R_{EMTrm}}{S_{Trm}} : \frac{\Sigma R_{MG}}{S_G}$

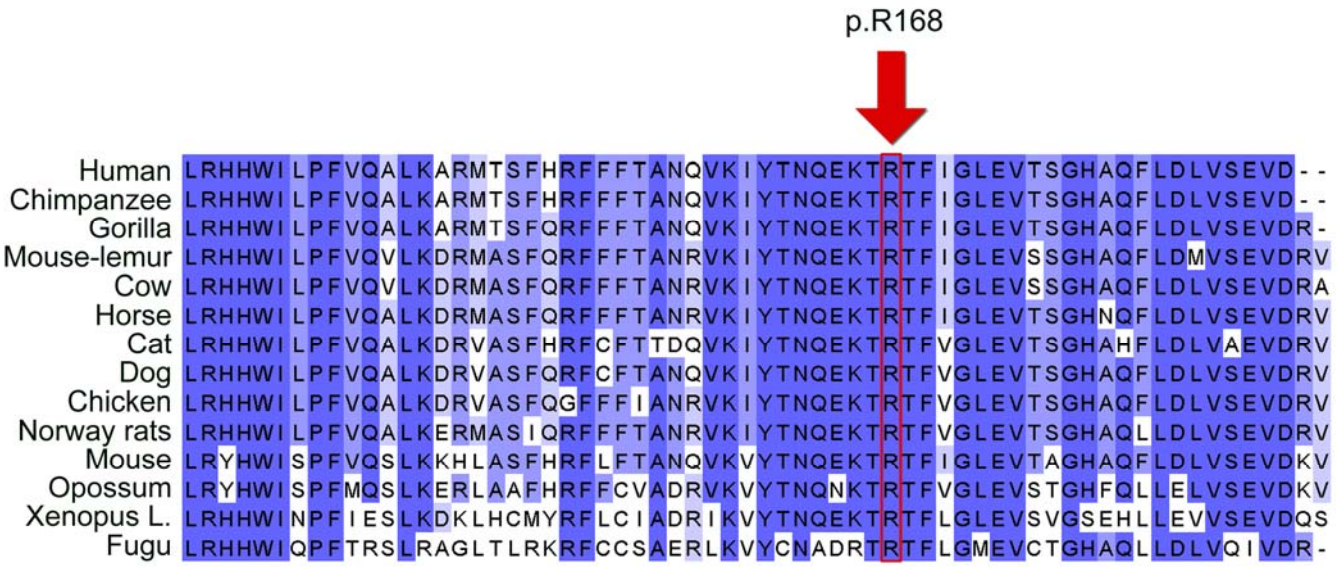
  R_{EMTrm}= reads equivalent mapped on the target region repeat-masked
  S_{Trm} = size of the target region repeat-masked
  R_{MG} = reads mapped on the genome (Eland results)
  S_G = size of the genome

Figure S2. Graphical View of Predicted Domain Structure of C16orf57 Protein



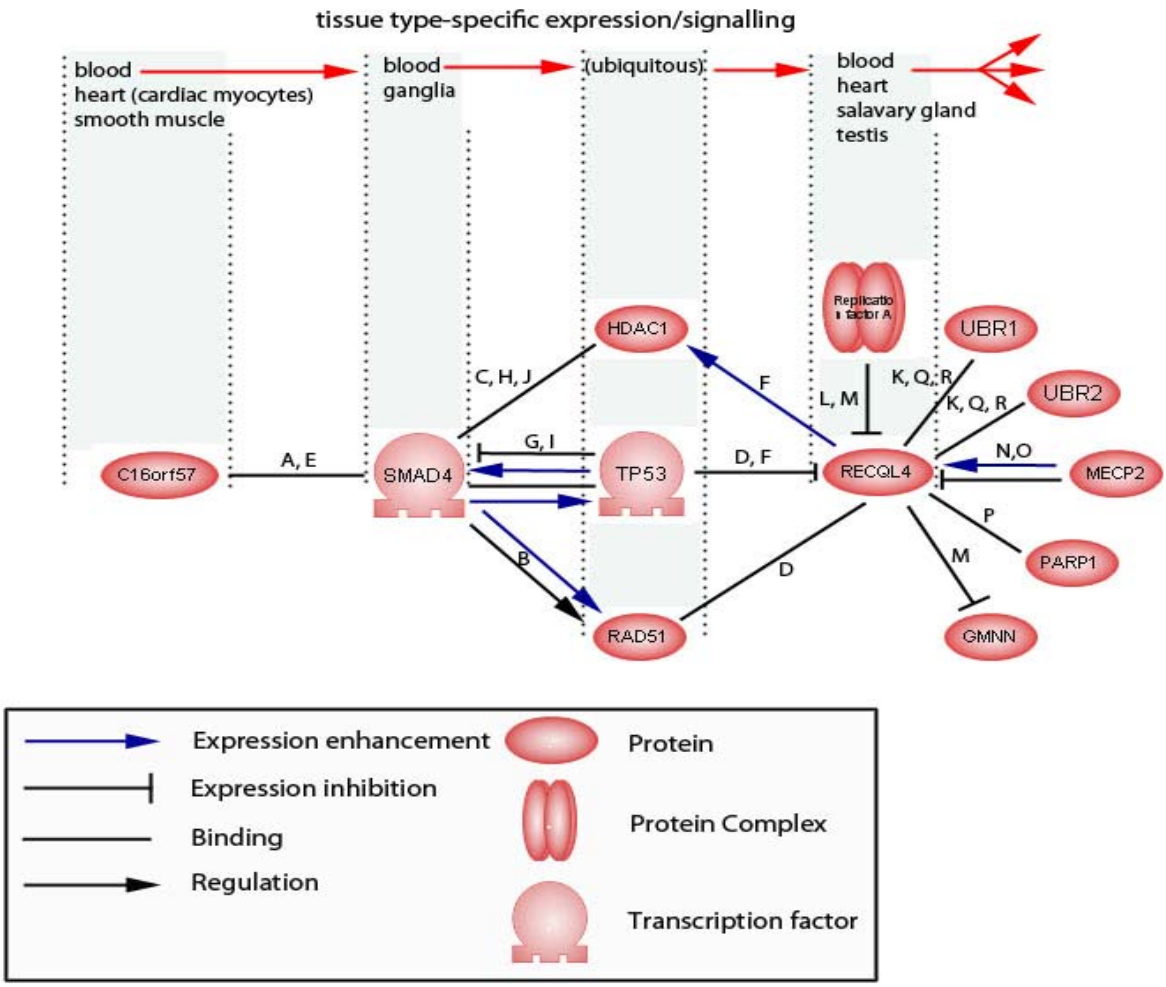
The green cylinders represent α helices, the yellow arrows β sheets and the thin black lines putative coils. The protein segments encoded by the seven *C16orf57* exons are indicated by red dashed lines. (A) Wild type C16orf57 protein (NP_078874). (B) C16orf57 protein conformation predicted by mutation c.504-2A>C present in the homozygous condition in the affected sibs of the inbred family. Upon skipping of exon 5 and frameshift, a truncated protein lacking 61 residues (boxed in purple) results (C), (D) Altered protein structures predicted by in frame loss of exon 4 and exon 6 encoded domains due to missense c.502A>G and c.666_676+1del12 mutations of the PN compound heterozygous patient. (PSIPRED Protein Structure Prediction Server)

Figure S3. C16orf57 Protein Conservation



Alignment of C16orf57 protein sequences from Human to Fugu by ClustalW. Identical aminoacid residues are in dark blue, similar residues in light blue. The arrow points to the functional, highly conserved and exposed p.R168 residue (ConSeq Server).

Figure S4. C16orf57 Taps into the RECQL4 Signalling Pathway through a Direct Interaction with SMAD4



Ref.A Colland, F., et al., Genome Res. 14: 1324-1332 (2004)
Ref.B Kanamoto, T., et al., EMBO J. 21 : 1219-1230 (2002)
Ref.C Kang, H.-Y., et al., J. Biol. Chem. 46 : 43749-43756 (2002)
Ref.D Petkovic, M., et al., J. Cell Sci. 118 : 4261-4269 (2005)
Ref.E Rual, J.F., et al., Nature 437: 1173-1178 (2005)
Ref.F Sengupta, S. et al., Oncogene 24: 1738-1748 (2005)
Ref.G Sorio, C., et al., Virchows Arch. 446: 239-245 (2005)
Ref.H Takeda, M., et al., Mol. Biol. Cel. 15 : 963-972 (2004)
Ref.I Venkatasubbarao, K., et al., Cancer Res. 65: 2861-2871 (2005)
Ref.J Wotton, D. et al., Cell 97 : 29-39 (1999)
Ref.K Yin, J., et al., Hum. Mol. Genet. 13 : 2421-2430 (2004)
Ref.L Macris MA, et al., DNA Repair (Amst) 5:172-180 (2006)
Ref.M Matsuno K, et al., Mol. Cell. Biol.26:4843-4852 (2006)
Ref.N Werner SR, et al., Biochem.Biophys.Res.Commun. 345:403-409 (2006)
Ref.O Jin W, et al., Hum. Genet. 123:643-653 (2008)
Ref.P Woo LL, et al., Exp. Cell Res. 312:3443-3457 (2006)
Ref.Q An JY, et al., Proc. Natl. Acad. Sci. U.S.A. 103:6212-6217 (2006)
Ref.R Tasaki T, et al., Mol. Cell. Biol. 25:7120-7136 (2005)

Two independent (protein interaction identification) studies have revealed direct interactions between the C16orf57 and SMAD4 proteins. This provides support for a model in which phenotypic overlap between RTS and PN can at least partially be explained by signalling of C16orf57 to RECQL4 through SMAD4 and HDAC1/TP53/RAD51. In addition, the fact that all these intermediate signalling proteins are expressed in blood (myeloid lineage), might explain the neutropenia as a distinctive phenotypic feature in patients with mutations in *C16orf57* (i.e. aberrant C16orf57 signalling). The model shown here was generated with the aid of Pathway Studio, whereas the interaction between C16orf57 and SMAD4 was identified by searching the STRING database and was originally identified both through Affinity Capture-MS assay (E) and a yeast-two-hybrid interaction trap (A).

A) Colland, F., Jacq, X., Trouplin, V., Mougin, C., Groizeleau, C., Hamburger, A., Meil, A., Wojcik, J., Legrain, P., Gauthier, J.M. (2004). Functional proteomics mapping of a human signaling pathway. Genome Res. 14, 1324-1332.

B) Kanamoto, T., Hellman, U., Heldin, C.H., Souchelnytskyi, S. (2002). Functional proteomics of transforming growth factor-beta1-stimulated Mv1Lu epithelial cells: Rad51 as a target of TGFbeta1-dependent regulation of DNA repair. EMBO J. 21, 1219-1230.

C) Kang, H.Y., Huang, K.E., Chang, S.Y., Ma, W.L., Lin, W.J., Chang, C. (2002). Differential modulation of androgen receptor-mediated transactivation by Smad3 and tumor suppressor Smad4. J. Biol. Chem. 277, 43749-43756.

D) Petkovic, M., Dietschy, T., Freire, R., Jiao, R., Stagljar, I. (2005). The human Rothmund-Thomson syndrome gene product, RECQL4, localizes to distinct nuclear foci that coincide with proteins involved in the maintenance of genome stability. J. Cell Sci. 118, 4261-4269.

E) Rual, J.F., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T., Dricot, A., Li, N., Berriz, G.F., Gibbons, F.D., Dreze, M., Ayivi-Guedehoussou, N., et al. (2005). Towards a proteome-scale map of the human protein-protein interaction network. Nature 437, 1173-1178.

F) Sengupta, S., Shimamoto, A., Koshiji, M., Pedeux, R., Rusin, M., Spillare, E.A., Shen, J.C., Huang, L.E., Lindor, N.M., Furuichi, Y., Harris, C.C. (2005). Oncogene 24, 1738-1748

G) Sorio, C., Capelli, P., Lissandrini, D., Moore, P.S., Balzarini, P., Falconi, M., Zamboni, G., Scarpa, A. (2005). Mucinous cystic carcinoma of the pancreas: a unique cell line and xenograft model of a preinvasive lesion. Virchows Arch. 446, 239-245.

H) Takeda, M., Mizuide, M., Oka, M., Watabe, T., Inoue, H., Suzuki, H., Fujita, T., Imamura, T., Miyazono, K., Miyazawa, K. (2004). Interaction with Smad4 is indispensable for suppression of BMP signaling by c-Ski. Mol. Biol. Cell. 15, 963-972.

I) Venkatasubbarao, K., Choudary, A., Freeman, J.W. (2005). Farnesyl transferase inhibitor (R115777)-induced inhibition of STAT3(Tyr705) phosphorylation in human pancreatic cancer cell lines require extracellular signal-regulated kinases. Cancer Res. 65, 2861-2871.

J) Wotton, D., Lo, R.S., Lee, S., Massagué, J. (1999). A Smad transcriptional corepressor. Cell. 97, 29-39.

K) Yin, J., Kwon, Y.T., Varshavsky, A., Wang, W. (2004). RECQL4, mutated in the Rothmund-Thomson and RAPADILINO syndromes, interacts with ubiquitin ligases UBR1 and UBR2 of the N-end rule pathway. Hum. Mol. Genet. 13, 2421-2430.

L) Macris, M.A., Krejci, L., Bussen, W., Shimamoto, A., Sung, P. (2006). Biochemical characterization of the RECQ4 protein, mutated in Rothmund-Thomson syndrome. DNA Repair (Amst) 5, 172-180.

M) Matsuno, K., Kumano, M., Kubota, Y., Hashimoto, Y., Takisawa, H. (2006). The N-terminal noncatalytic region of Xenopus RecQ4 is required for chromatin binding of DNA polymerase alpha in the initiation of DNA replication. Mol. Cell. Biol. 26, 4843-4852.

N) Werner, S.R., Prahalad, A.K., Yang, J., Hock, J.M. (2006). RECQL4-deficient cells are hypersensitive to oxidative stress/damage: Insights for osteosarcoma prevalence and heterogeneity in Rothmund-Thomson syndrome. Biochem. Biophys. Res. Commun. 345, 403-409.

O) Jin, W., Liu, H., Zhang, Y., Otta, S.K., Plon, S.E., Wang, L.L. (2008). Sensitivity of RECQL4-deficient fibroblasts from Rothmund-Thomson syndrome patients to genotoxic agents. Hum. Genet. 123, 643-653.

P) Woo, L.L., Futami, K., Shimamoto, A., Furuichi, Y., Frank, K.M. (2006). The Rothmund-Thomson gene product RECQL4 localizes to the nucleolus in response to oxidative stress. Exp. Cell Res. 312, 3443-3457.

Q) An, J.Y., Seo, J.W., Tasaki, T., Lee, M.J., Varshavsky, A., Kwon, Y.T. (2006). Impaired neurogenesis and cardiovascular development in mice lacking the E3 ubiquitin ligases UBR1 and UBR2 of the N-end rule pathway. Proc. Natl. Acad. Sci. U. S. A. 103, 6212-6217.

R) Tasaki, T., Mulder, L.C., Iwamatsu, A., Lee, M.J., Davydov, I.V., Varshavsky, A., Muesing, M., Kwon, Y.T. (2005). A family of mammalian E3 ubiquitin ligases that contain the UBR box motif and recognize N-degrons. Mol. Cell. Biol. 25, 7120-7136.

Table S1. Results of Genome-wide Linkage Analysis Using 262K NspI SNP Array (Affymetrix) Platform, Showing Three Regions with LOD Score >2.5

Chr.	Cytoband	Max LOD score	SNP Max LOD score	Interval size (Mbp)	Gene content
2	q22.2	2.6	SNP_A-4208377	0.04	*KYNU, ARHGAP15*
6	p22.1	2.8	SNP_A-1894559	0.15	*HLA-G*
16	q12.2-q21	2.93	SNP_A-1804307	3.4	> 80 genes

Table S2. Ranking of Candidate Homozygous Mismatches According to Location and Evolutionary Conservation

Base change	Genomic position	Gene	Location	Mammal conserv.
A>C	**56608737**	***C16orf57***	**acceptor splice site**	**very high**
C>T	55415773	*NUP93*	intron	high
A>G	55742632	*CPNE2-NIP30*	near gene (2kb)	high
C>T	55848191	*PLLP*	3' UTR	low
G>C	54698403	*BC035113*	intron	moderate
A>G	55107654	*BBS2*	intron	moderate
G>A	55738779	*CPNE2*	intron	moderate
G>A	56222743	*GPR56*	intron	moderate
C>G	57061773	*NDRG4*	intron	moderate
T>C	54459899	*CES7*	intron	low
C>T	54876398	*GNAO1*	intron	low
C>T	54941945	*GNAO1*	intron	low
A>G	55586259	*NLRC5*	intron	low
G>A	55766270	*NIP30*	intron	low
G>A	55795497	*RSPRY1*	intron	low
C>T	56045488	*CoQ9*	intron	low
T>C	57085085	*NDRG4*	intron	low

Table S3. Primer Pairs and PCR Conditions for *C16orf57* Genomic Analysis

C16orf57	Primers	Sequence	A.T. (°C)	Product size (bp)
Exon 1	Forward	AGTCGGTGGGCTGAATCT	60	325
	Reverse	TGGCTCTAGGGTGAATGC		
Exon 2	Forward	ACACACACTCAGAGCCACCA	60	405
	Reverse	AATGACTTTCCCACCACCAG		
Exon 3	Forward	CCTTCTGGGCTTCTTCAT	56	455
	Reverse	CATCAGGGGTGTCAAGAG		
Exon 4	Forward	ACCTGGATGATGTTGTGTGT	56.8	461
	Reverse	CTACTGTGCCTGGGATCT		
Exon 5	Forward	GGAGCAGGAAAGCGAGTGTA	60	286
	Reverse	ATGGGTCAATGGAGAAGCAG		
Exon 6	Forward	TTGTCTGTCTGTGGAGGGTG	60	352
	Reverse	CAGGGATTCAGGTTTGGGTG		
Exon 7	Forward	AGCATCTGTGTCCTCATCTG	59	374
	Reverse	GTTCCTCCATCTCAGCCTG		

A. T. = Annealing temperature

Table S4. Primer Pairs and PCR Conditions for *C16orf57* Transcripts Analysis of the (V-2) Homozygous and (P) Compound Heterozygous PN Patients

Primer position	Primers	Sequence	A. T. (°C)	Product size (bp)		See results (Fig.)
				exp.	obs.	
exon 4 3'UTR	Forward	ACACCAATCAAGAGAAAACC	56	374	268	2C
	Reverse	GTTCCTCCATCTCAGCCTG				
exon 3/ exon 4 3'UTR	Forward	CCTCCTTCCACAGATTCTTC	55	419	335	2F
	Reverse	GTTCCTCCATCTCAGCCTG				
exon 3 exon 5/ exon 6	Forward	TCTGAAAGCCCGTATGACC	58	199	145	2H
	Reverse	AGAAGGATCCTGGTAGAAAGTG				

A. T. = Annealing temperature