The American Journal of Human Genetics, Volume *92*

Supplemental Data

FAM111A Mutations Result in Hypoparathyroidism and Impaired Skeletal Development

Sheila Unger, Maria W. Górna, Antony Le Béchec, Sonia Do Vale-Pereira, Maria Francesca Bedeschi, Stefan Geiberger, Giedre Grigelioniene, Eva Horemuzova, Faustina Lalatta, Ekkehart Lausch, Cinzia Magnani, Sheela Nampoothiri, Gen Nishimura, Duccio Petrella, Francisca Rojas-Ringeling, Akari Utsunomiya, Bernhard Zabel, Sylvain Pradervand, Keith Harshman, Belinda Campos-Xavier, Luisa Bonafé, Giulio Superti-Furga, Brian Stevenson, and Andrea Superti-Furga

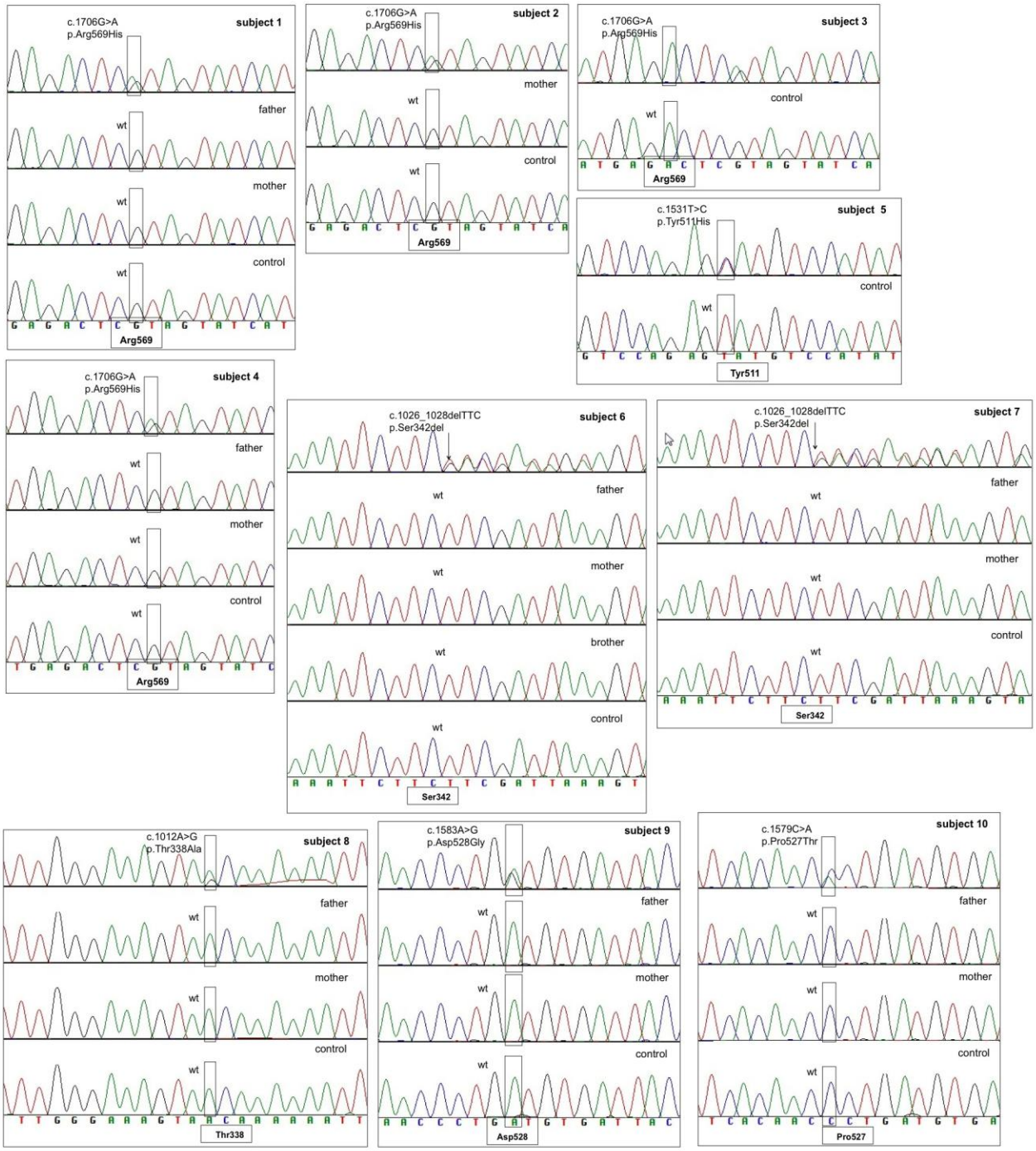

Figure S1. Electropherograms of Sanger Sequencing Reactions of *FAM111A* Mutations in Affected Individuals and Family Members

Sanger sequencing of *FAM111A* (RefSeq: NM_001142519.1, NP_001135991.1; Ensembl: gene ENSG00000166801; transcript ENST00000361723) amplicons from DNA extracted from blood cells, saliva, and cultured fibroblasts was performed using routine methods as described before [1]; primer sequences to amplify the exons and exon/intron junctions of *FAM111A* are in Supplementary Table 3, below. All mutations identified were double-checked with bidirectional sequences on fresh PCR products.

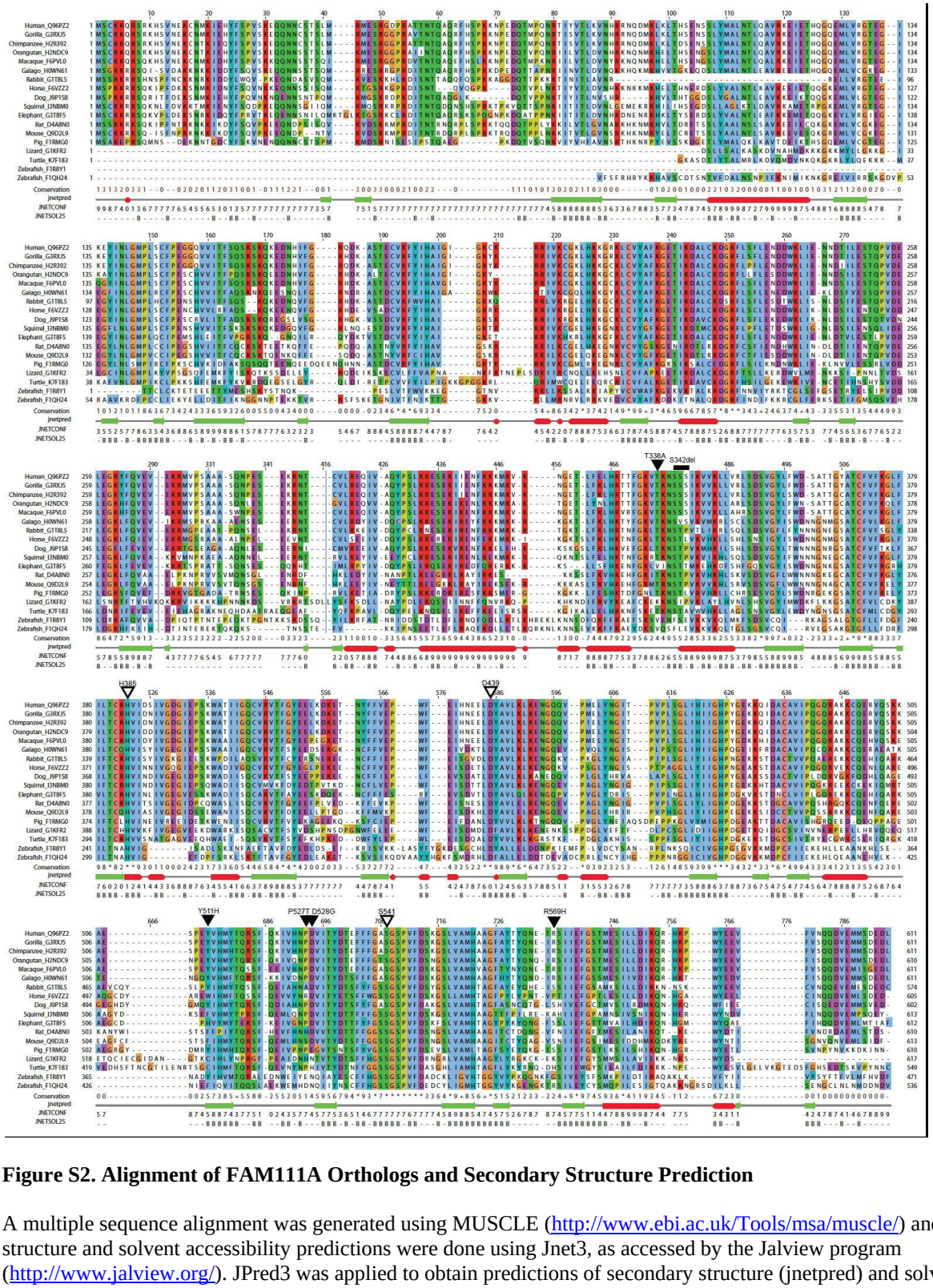

Figure S2. Alignment of FAM111A Orthologs and Secondary Structure Prediction

A multiple sequence alignment was generated using MUSCLE (http://www.ebi.ac.uk/Tools/msa/muscle/) and secondary structure and solvent accessibility predictions were done using Jnet3, as accessed by the Jalview program (http://www.jalview.org/). JPred3 was applied to obtain predictions of secondary structure (jnetpred) and solvent accessibility for the human FAM111A (http://www.compbio.dundee.ac.uk/www-jpred/). In the JNETSOL25 line, "B" indicates a residue predicted to be buried with less than 25% solvent accessibility. Annotated are the UniProt identifiers of the protein sequences, the confidence score of the secondary structure prediction (JNETCONF), as well as the conservation score per residue. The empty triangles indicate the conserved putative catalytic residues, and black triangles indicate mutations found in human *FAM111A*. The three-serine stretch around S342 is indicated with a black line. Vertical blue lines indicate where long linker sequences (residues 178-203 and 230-294) in the Lizard *Fam111a* were omitted from the figure.

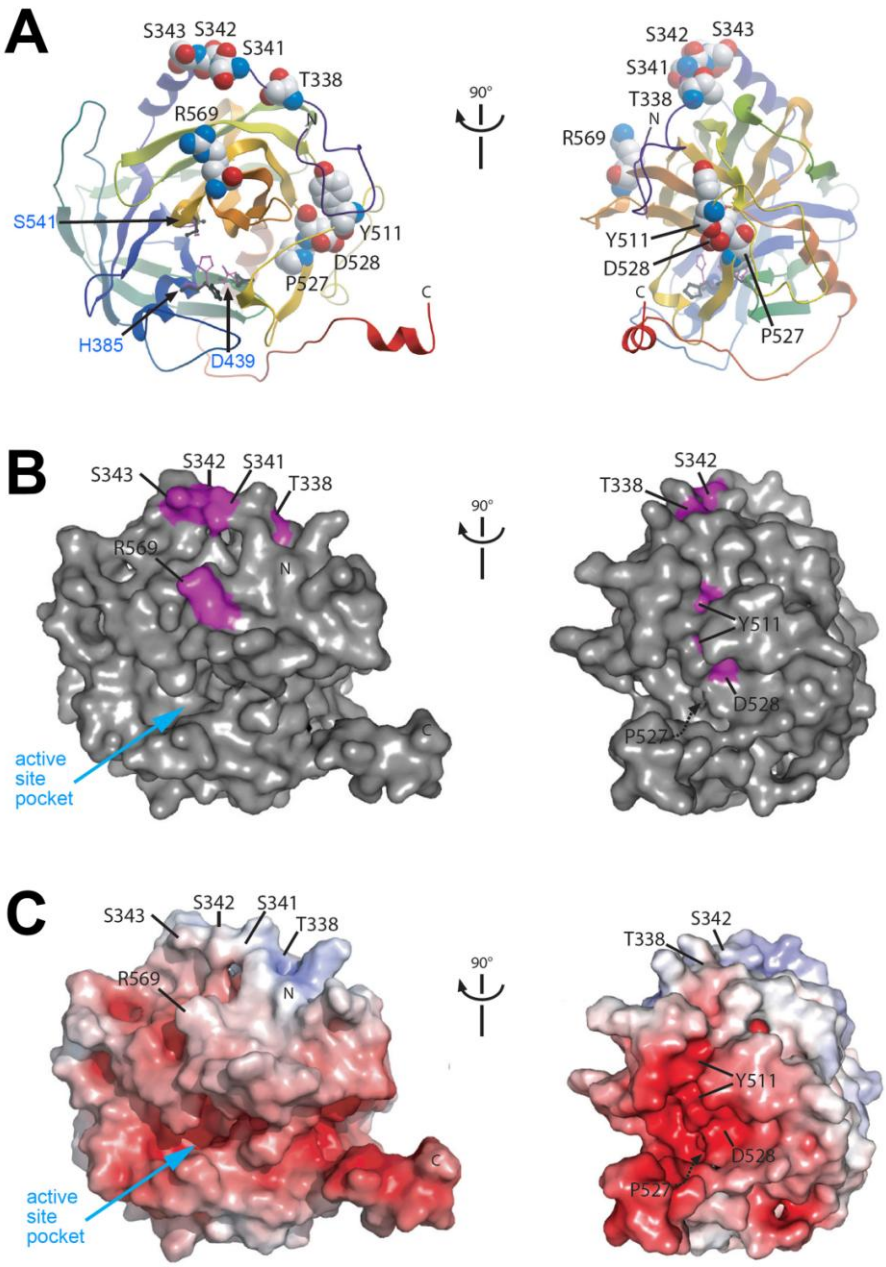

Figure S3. Molecular Modeling of the Carboxy-Terminal, Putative Peptidase Domain of FAM111A

The comparative model of FAM111A was obtained using the I-TASSER server [2; 3](http://zhanglab.ccmb.med.umich.edu/I-TASSER/). FAM111A shares 22% sequence identity in the threading aligned region with the top template (PDB entry 3qo6). The surface electrostatic charge was calculated using the APBS PyMOL plugin (http://apbs.sourceforge.net). ICM Browser (MolSoft; www.molsoft.com/icm_browser.html) and PyMOL (DeLano Scientific, now Schrödinger K.K.; www.pymol.org/) were used for figure preparations. - A: The comparative model of the C-terminal region of human FAM111A (residues 322-611) is shown in cartoon representation. The residues labeled in blue (H385, D439 and S541) are the so-called catalytic triad. Also shown is the catalytic triad of the template protease Deg1 (PDB entry 3qo6; purple sticks). The residues mutated in KCS/OCS are depicted as spheres. Note that p.Ser342del occurs in a stretch of three serines; at the protein level, it is irrelevant to distinguish between the three residues. B: surface representation of the model of the FAM111A protease-like domain. The residues affected by the mutations are highlighted in pink. C: The surface electrostatic charge is displayed on the same model (negative charge in red, positive charge in blue). Note the negative charges around the active site pocket. - Panels C and D show that the residues mutated in KCS/OCS are not close to the active site pocket, but are instead clustered in a segment of the globular protein (all residues are visible in the two adjacent views rotated by 90 degrees) and are at or just beneath to the surface of the protein.

Table S1. Statistics of Whole Exome Sequencing Experiments in 1 Individual with KCS (Individual 1) and Three Individuals with OCS (Individuals 8, 9, and 10)

	Individual 1 (KCS)	Individual 8 (OCS)	Individual 9 (OCS)	Individual 10 (OCS)
Total number of sequenced reads	80'361'230	82'789'860	83'102'584	83'341'110
Total number of unique reads	66'473'866	66'927'820	72'458'912	73'930'973
Total number of unique reads mapped	65'350'394	65'769'943	71'239'927	72'715'713
Total number of bases mapped (Gb)	6.58	6.63	7.18	7.33
Total bases mapping to targets (Gb)	5.73	5.76	6.15	6.34
% Targets with 10× coverage	98.06	97.92	98.26	98.18
Mean target coverage	91.34	91.4	98.8	102.1

For library construction, 3 µg of genomic DNA were fragmented with a Covaris S2 sonicator (Covaris, Inc.) to obtain DNA fragments with a homogenous size distribution having a maximum at 180-200 base pairs. Fragmented DNA was purified with AMPure XP beads and the quality of the fragmented DNA was assessed with an Agilent Bioanalyzer. Preparation of the exome enriched, barcoded sequencing libraries was perfomed using Agilent SureSelect Human All Exon v3 kit. The final libraries were quantified with a Qubit Fluorometer (Life Technologies) and the correct size distribution was validated with an Agilent Bioanalyzer. Libraries were sequenced on Illumina HiSeq 2000, generating 100 bp paired-end reads. Purity filtered reads were aligned to the human genome (assembly hg19) with the Burrows-Wheeler Aligner version 1.5.9 using default parameters (BWA) [4].

Table S2. Data Analysis Pipeline Following Whole-Exome Sequencing

	Individual 1 (KCS)	Individual 8 (OCS)	Individual 9 (OCS)	Individual 10 (OCS)
High-confidence variant calls	509'720	558'451	626'038	562'611
After exclusion of nongenic, UTR and intronic variants	25'738	26'458	25'857	25'825
After exclusion of synonymous variants	12'969	13'418	13'060	12'957
After exclusion of known variants	160	234	143	285
Fit dominant model of inheritance	144	213	128	256
Number of genes mutated in all 4 patients	1	1	1	1

Aligned data were processed with the Picard pipeline to mark duplicated reads (http://picard.sourceforge.net/). The Genome Analysis Toolkit version 1.6.5 (GATK) [5] was used for (1) local realignment around indels, (2) base quality score recalibration, (3) single-nucleotide variants (SNVs) and indels discovery, and (4) variants quality score recalibration (VQSR) following GATK best practices recommendations (http://www.broadinstitute.org/gatk/). Variants databases used were dbSNP 135, HapMap 3.3 and the 1000 genomes Omni 2.5. Variants were annotated using ANNOVAR [6], filtered by the presence in dbSNP [7], 1000 genomes [8], 6500NHLBI (http://evs.gs.washington.edu/EVS/), the CoLaus cohort [9] and our internal database, and prioritized according to putative functionality (missense and nonsense).

Table S3. Primers and PCR Conditions Used to Amplify the Coding Region of *FAM111A*

Primer	Sequence 5'-3'	Product Size (bp)	PCR Conditions
FAM111A_Ex1F	AGC AAG GTT GGA GCC TAA GG	331	62
FAM111A_Ex1R	GCA TTC ATG GAT TTT GGT ATC C		
FAM111A_Ex2Fa	TCA CTG GCC CTT TTA GAT TGG	1035*	Touchdown
FAM111A_Ex2Ra	TCT ATG CAA TTC AAA TAA TGT TTC C		
FAM111A_Ex2Fb	AGC CAG TTG ATG AAT TAG AAG G	473	Touchdown
FAM111A_Ex2Rb	TGT CAC CCT TAC ACA TTG ACC		
FAM111A_Ex2Fc	AGT TGG GTA CTT ATT CTG GGA C	431	62
FAM111A_Ex2Rc	GTT CCT GAC ATT TCT TTG CTC G		
FAM111A_Ex2Fd	GTA CCT ATG GAA CTA TAT AAT GG	542	Touchdown
FAM111A_Ex2Rd	AAT GCC TAT GAA ATA ACA ACT CC		
***For sequencing an additional internal primer was used: FAM111A_Ex2FaSeq: GTA CAT AAA CCT TGG AAT GCC**			

Standard PCR conditions were: initial denaturation at 95°C for 10min followed by thirty-five cycles of denaturation at 95°c for 30s, annealing at 62°C for 1min, and extension at 72°C for 1min30s with final extension for 10min at 72°C. For Touchdown PCR, the conditions were: initial denaturation at 95°C for 10min followed by fifteen touchdown cycles of denaturation at 95°c for 30s, annealing at 68°C for 30s, and extension at 72°c for 1min30s; then thirty cycles of denaturation at 95°c for 30s, annealing at 53°C for 30s, and extension at 72°c for 1min30s with final extension for 10min at 72°C.

References

1. Campos-Xavier, A.B., Martinet, D., Bateman, J., Belluoccio, D., Rowley, L., Tan, T.Y., Baxova, A., Gustavson, K.H., Borochowitz, Z.U., Innes, A.M., et al. (2009). Mutations in the heparan-sulfate proteoglycan glypican 6 (GPC6) impair endochondral ossification and cause recessive omodysplasia. American Journal of Human Genetics 84, 760-770.

2. Roy, A., Kucukural, A., and Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. Nature Protocols 5, 725-738.

3. Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. BMC Bioinformatics 9, 40.

4. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754-1760.

5. DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature Genetics 43, 491-498.

6. Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Research 38, e164.

7. Sherry, S.T., Ward, M.H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M., and Sirotkin, K. (2001). dbSNP: the NCBI database of genetic variation. Nucleic Acids Research 29, 308-311.

8. 1000 Genomes Project, C., Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., and McVean, G.A. (2012). An integrated map of genetic variation from 1,092 human genomes. Nature 491, 56-65.

9. Firmann, M., Mayor, V., Vidal, P.M., Bochud, M., Pecoud, A., Hayoz, D., Paccaud, F., Preisig, M., Song, K.S., Yuan, X., et al. (2008). The CoLaus study: a population-based study to investigate the epidemiology and genetic determinants of cardiovascular risk factors and metabolic syndrome. BMC Cardiovascular Disorders 8, 6.