Praktisches Handling von NGS Daten

Saskia Biskup, Dr. med. Dr. rer. nat.

Hertie Institut für Klinische Hirnforschung
Institut für Klinische Genetik, Klinikum Stuttgart
Praxis für Humangenetik / CeGaT GmbH

Aspekte

Forschung:

- ***Entdeckung neuer Varianten (Krankheitsrelevanz ?)***
- ***Mutational Load bei verschiedenen Erkrankungen***
- ***funktionelle Untersuchungen***

Klinik:

- ***Bedarf aus klinischen Fragestellungen***

  → Rundum-Sorglos-Paket für den Kliniker

Qualität:

- ***Gut ist nicht gut genug: “Enrichment” + “Auswertung”***
- ***Standardisierung / Automation / Akkreditierung NGS***

CeGaT business units

CeGaT GmbH	Scope		
Business Unit 1: Diagnostic Panels	▪ Dementia / ALS ▪ Parkinson ▪ Epilepsy	▪ Hereditary eye diseases ▪ Neuromuscular diseases ▪ Pharmacogenomic Panel	▪ Hereditary Cancer / Tumor ▪ Hereditary Deafness
Business Unit 2: Molecular Diagnostics	▪ Dementia / ALS ▪ Parkinson ▪ Dystonia ▪ Epilepsy	▪ Hereditary eye diseases ▪ Neuropathy ▪ Rare diseases ▪ Metabolic Disorders	▪ Hereditary Cancer / Tumor
Business Unit 3: NGS[1]	▪ Genome ▪ Exome ▪ Transcriptome	▪ Metagenome ▪ microRNA	

Cooperation: FISH[2]	▪ Identification of:
	  ▪ novel amplified genes
	  ▪ deleted genes or
	  ▪ translocated genes
		▪ Visualization of genes and their qualitative / quantitative changes
		▪ Individual development of specific FISH-Assays	

[1] NGS: Next-Generation Sequencing

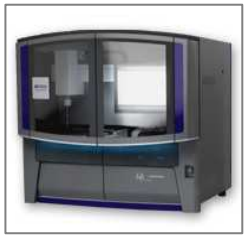
[2] FISH: Fluorescence in-situ hybridisation, in cooperation together with Prof. Perner, Director of the Institute of Prostate Cancer Research, University Hospital of Bonn

Gene Sequencing – Different approaches for different demands

CeGaT Center for Genomics and Transcriptomics

Sanger Sequencing	Panel Sequencing	Exome Sequencing
⊕ Established method	⊕ Parallel sequencing of all genes associated with a certain disease	⊕ Parallel sequencing of all coding regions
⊖ Sequencing of single genes only	⊕ Time and cost efficient	⊖ High complexity
⊖ Time consuming &	⊕ High chance to identify the genetic cause of the disease	
⊖ expensive		
• Little chance to identify the underlying genetic cause		

www.lifetechnologies.com



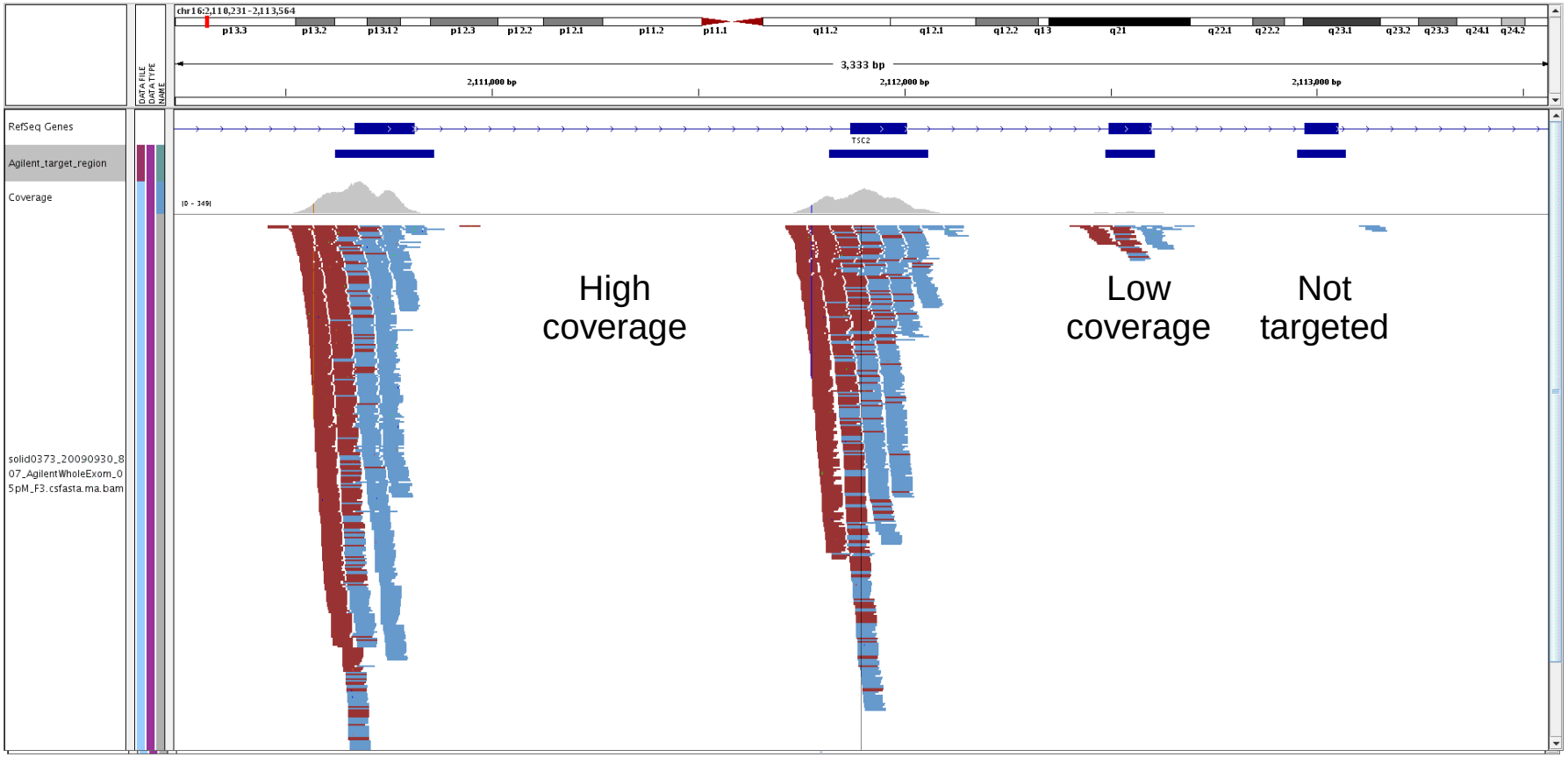
www.lifetechnologies.com

www.nih.gov

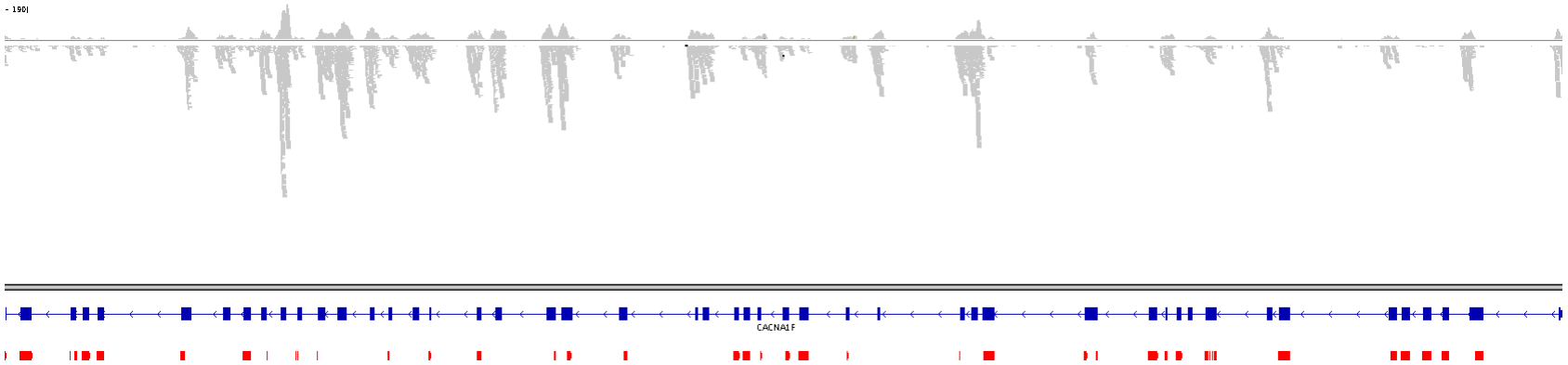
CeGaT Panel for Neurodegenerative Diseases 277 – includes 16 Subpanels / Candidate Genes / GWAS loci

Subpanels for Diagnostic testing	
3 Panels for **Parkinson´s Disease (autosomal dominant, recessive, atypical)**	20 genes
2 Panels for **Alzheimer Dementia** & **Frontotemporal Dementia**	13 genes
1 Panel for **Amyotrophic Lateral Sclerosis**	20 genes
4 Panels for **Dystonia**	11 genes
1 Panel for **Neuroakanthocytosis**	4 genes
1 Panel for **Neurdegeneration with brain iron accumulation**	8 genes
1 Panel for **Ceroidlipofuscinosis**	9 genes
1 Panel for **Leukodystrophies**	20 genes
1 Panel for **CMT**	55 genes
1 Panel for **HSP**	110 genes

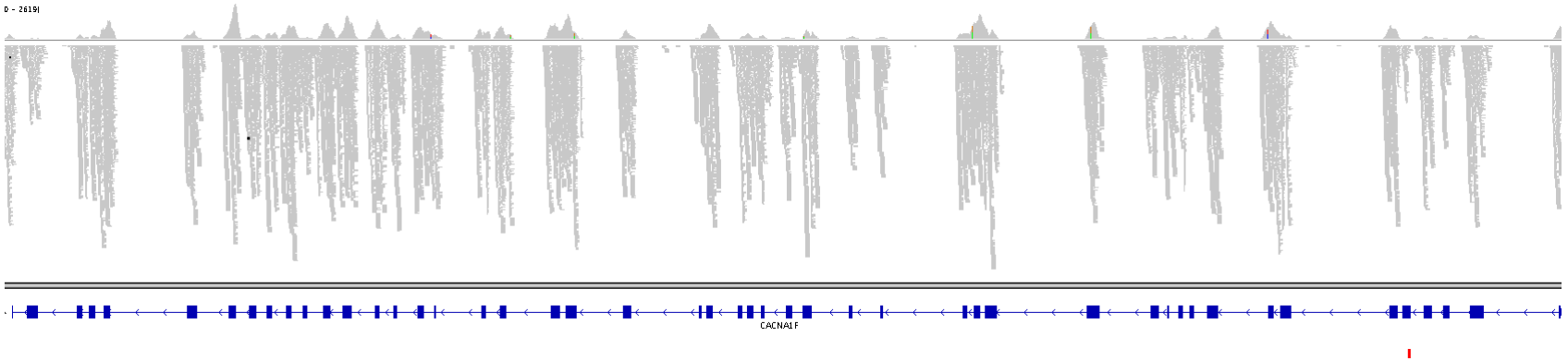
Issues regarding Uniformity and missed regions Targeted resequencing



Coverage Exome



Coverage Panel



The CLARITY Challenge (Children´s Hospital Boston)

- Genomes / exomes in a diagnostic setting
- Standardizing NGS analysis (filtering steps, prediction tools)
- Interpretation and reporting to clinicians and patients
- How far are we away?

The CLARITY Challenge – 30 Teams competing

- BGI (Shenzhen, China)
- Brigham and Women's Hospital, Division of Genetics (Boston, Massachusetts)
- British Columbia Cancer Agency (Vancouver, Canada)
- Children's Hospital of Eastern Ontario (Ottawa, Canada)
- Clinical Institute of Medical Genetics (Ljubljana, Slovenia)
- Genedata AG (Basel, Switzerland)
- CeGaT (Tübingen, Germany), Institute of Pathology, Bonn, Genomatix (Munich)
- Genome Institute of Singapore Agency for Science, Technology and Research (A*STAR) (Biopolis, Singapore)
- HudsonAlpha Institute for Biotechnology (Huntsville, Alabama)
- Institute for Systems Biology (Seattle, Washington)
- IRCCS Casa Sollievo della Sofferenza (San Giovanni Rotondo, Foggia, Italy)
- National Institutes of Health (Bethesda, Maryland)
- NextBio (Santa Clara, California)
- Omicia Inc/University of Utah (supported by LocusDev Inc) (Emeryville, California)
- Pearlgen (Chapel Hill, North Carolina)
- Radboud University Nijmegen Medical Center (Nijmegen, Netherlands)

The CLARITY Challenge – 30 Teams competing

- Sanofi (Cambridge, Massachusetts)
- Science For Life Laboratory (SciLifeLab), Karolinska Institute (Solna, Sweden)
- Scripps Genomic Medicine, Scripps Translational Science Institute(San Diego, California)
- Seven Bridges Genomics (Cambridge, Massachusetts)
- SimulConsult / Geisinger (Chestnut Hill, Massachusetts / Danville, Pennsylvania)
- SNPedia (Potomac, Maryland)
- Strand Life Sciences (Bangalore, India)
- Tel Aviv University (Israel)
- The Medical College of Wisconsin (Milwaukee, Wisconsin)
- The Research Institute at Nationwide Children's Hospital (Columbus, Ohio)
- The University of Texas Health Science Center at Houston, Universidad de Cantabria (Santander, Spain)
- University of Iowa (Iowa City, Iowa)
- Yale School of Public Health, Division of Biostatistics (New Haven, Connecticut)

Nature Medicine |
07 Nov 2012 | 12:32 EST | Posted by Susan Matthews | Category: Genetics

…
In January, the Children's Hospital put out the call for submissions, asking participants to help determine the unknown genetic root cause of illness in three children. The teams could sequence the genomes of the children and their parents, and were tasked with interpreting the information. The ultimate aim of the competition was to shed light on how data from whole genome sequences can be made most useful in a clinical setting.

There was a "real question of whether these technologies are ready for prime-time clinical applications," says Isaac Kohane, an endocrinologist at Children's Hospital. "What these teams have demonstrated is that going from end to end—from a genome sequence to a clinical readable report—can be turned into a routine process."

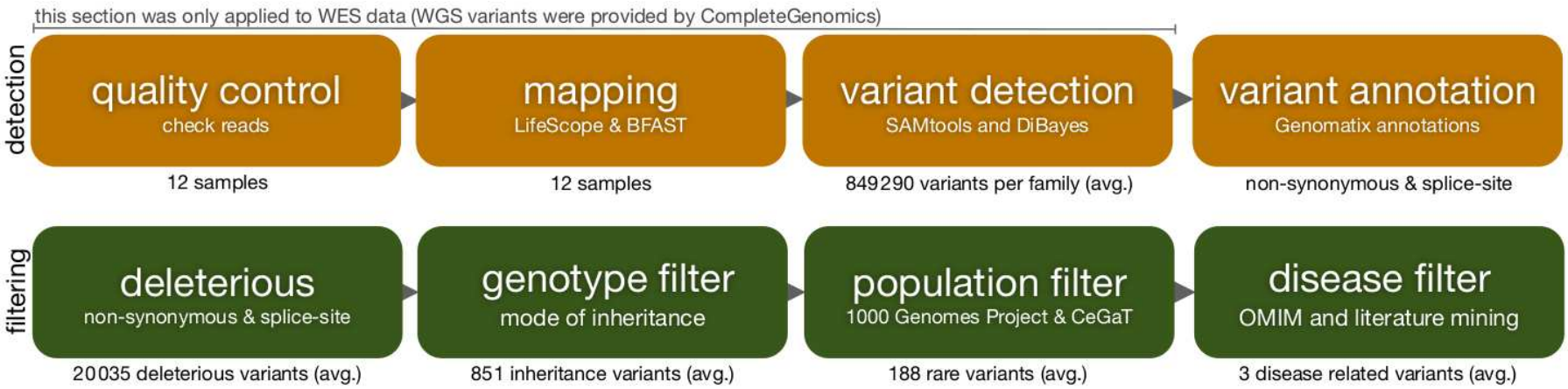
.. A German team (with representatives from the gene sequencing companies Genomatix and CeGaT, as well as the Institute of Pathology at the University of Bonn) also received $5,000 as a finalist for flagging all likely genetic mutations in the three cases.

Whole exome sequencing (WES)

12 DNA samples were received at the LIFE lab in LIFE/Carlsbad for initial processing.They were then sheared and made into SOLiD 5500 fragment libraries as per standard protocol for the LIFE Library Builder (further details where provided by Children's). From this dataset Children's provided all xsq-files and .BAM files from the mapping.

Whole genome sequencing (WGS)

Whole genome sequencing was performed by CompleteGenomics for 10 out of the 12 family members (see Table 1). Children's provided all sequencing, mapping and variant calling data generated by CompleteGenomics.



Sample	SNVs	INDELs	non-synonymous	Ti/Tv	hom	het	ref
W1-1	215565	7467	11966	2.3	144825	78207	32464
W1-2	192344	6733	11722	2.3	129701	69376	31934
W1-3	203541	7051	12025	2.4	135987	74605	32913
W2-1	200564	6953	10329	2.3	116995	90522	77601
W2-2	143170	5424	9963	2.3	85373	63221	69614
W2-3	169479	6075	10003	2.3	99638	75916	77014
W2-4	146418	5599	9975	2.3	87710	64307	69624
W2-5	127835	4960	9768	2.3	76683	56112	67955
W2-6	151224	5513	10008	2.3	91774	64963	70570
W3-1	141914	5444	11945	2.4	89373	57985	27762
W3-2	135569	5031	11795	2.4	84796	55804	28431
W3-3	138549	4896	11932	2.4	89535	53910	27209

Table 2: Variant calling statistics. For each sample the number of called variants with SAMtools is listed. Non-synonymous variants are those that affect the protein sequence. The Ti/Tv column specified the transition/transversion ratio. The last three columns specify the genotype of the SNP. In case of ref, a variant called in an other family member was detected as reference in the sample.

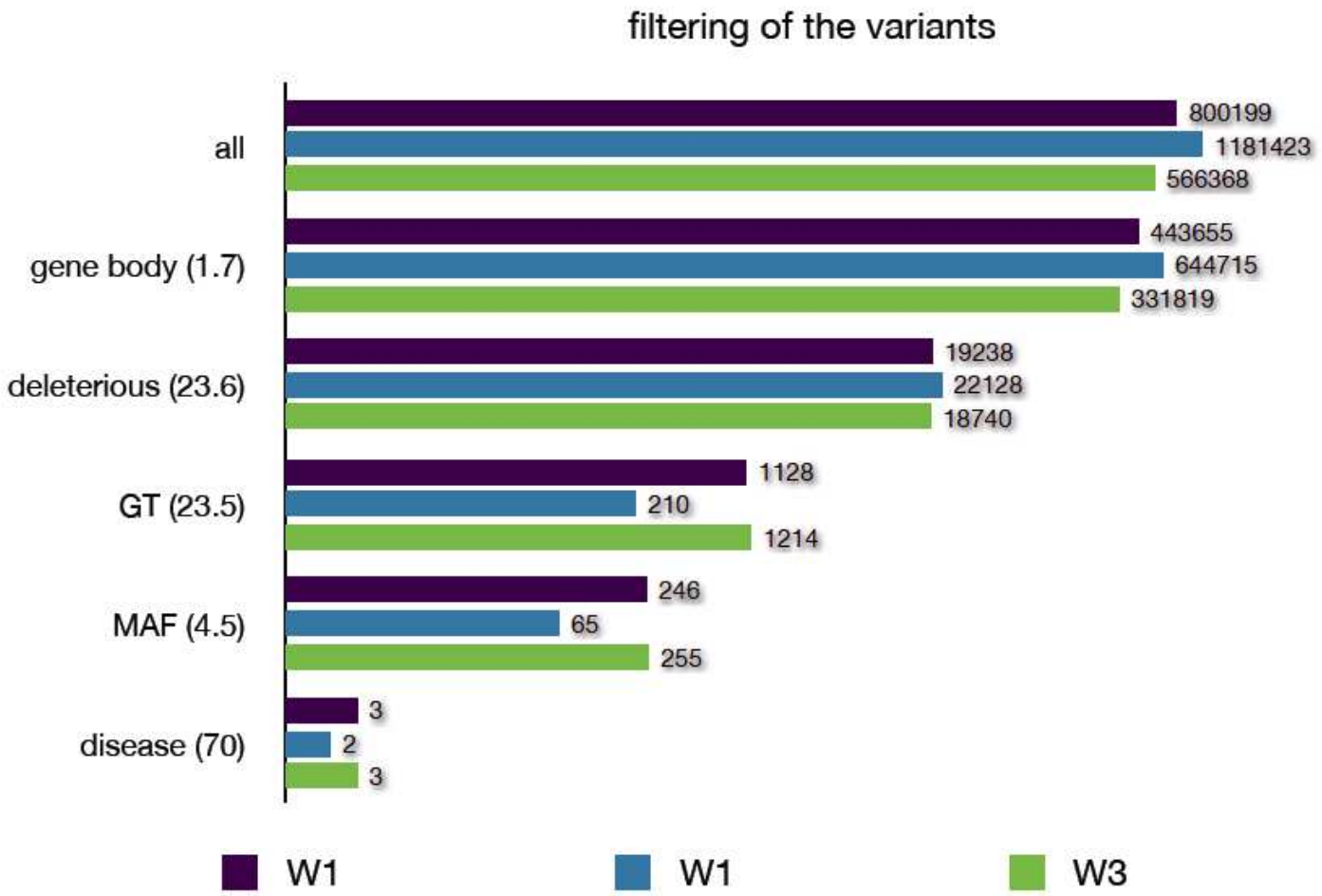


Figure 1: Filtering of variants for W1, W2 and W3 (log scale). The first bars (all) show the union of variants called in any of the family members. For each subsequent filtering step the number of remaining variants and the filter factor is given. The filters are: gene body (variants that overlap with a transcript), deleterious (variants that alter the protein sequence or hit a canonical splice-site), GT (genotype filter derived from the medical report), MAF (1000 Genomes Project background filter), disease (filter for the primary disease and MESH parents from the medical report).

Family 1: Centronuclear Myopathy