Discover putative disease causing mutations in patients with ataxia or paraplegia by using Next -Generation -Sequencing panels

Martin Schulze (1) and Florian Harmuth (1), Julia Bickmann (1), Marc Sturm (1), Julia Schicks (2), Matthis Synofzik (2), Ebba Lohmann (3), Olaf Riess (1), Ludgar Schöls (2) and Peter Bauer (1)

(1) Institute of Medical Genetics and Applied Genomics, Eberhard-Karls-University, Tuebingen, Germany
(2) Department of Neurology and Hertie Institute for Clinical Brain Research, German Center of Neurodegenerative Diseases, Eberhard-Karls-University, Tuebingen, Germany
(3) Department of Neurology, Istanbul Medical Faculty, Istanbul University, Istanbul, Turkey

Summary

To date, only a subset of genes known to be involved in Mendelian causes of ataxia and paraplegia have been systematically tested in cohorts of patients. Within NeurOmics work package 5 (WP5) we now investigate pre-screened patient cohorts for the occurence of putative disease-causing variants. Therefore, we established separate ataxia and paraplegia gene panels (HaloPlex, Agilent) to specifically enrich DNA target regions followed by next-generation-sequencing (NGS). Sequenced reads were analysed with our in-house bioinformatics pipeline. We found several rare, potentially disease-causing variants in different genes (ataxia: *CACNA1A*, *FGF14* / paraplegia: *SPG11*, *KIF1C*) covered by more than 20 reads which is diagnostically relevant.

1 Introduction

Within the last decade, knowledge about genetics of neurodegenerative diseases such as hereditary ataxia and paraplegia could greatly be expanded. Diseases with ataxia and paraplegia are highly heterogeneous and show an autosomal dominant, autosomal recessive or X-linked mode of inheritance. To date, more than 100 genes have been identified to be involved in Mendelian causes of ataxia and paraplegia. Despite this knowledge, only a subset of these genes has been tested systematically for potential disease-causing mutations in patient cohorts.
Within WP5 of the NeurOmics project, it is planned to sequence 100 pre -screened patients of each disease group from families with at least two affected members with unknown genetic cause using gene panels. Therefore, we have established separate ataxia and paraplegia specific selector probe based enrichment assays (HaloPlex, Agilent).
Here we show first results of subsets of 28 ataxia samples and 9 paraplegia samples.

2 Objective

Wp5 of the NeurOmics project aims to identify putative disease-causing variants by using separate ataxia and paraplegia specific gene panels (HaloPlex, Agilent) in combination with NGS sequencing (MiSeq, Illumina).

3 HaloPlex panels (design)

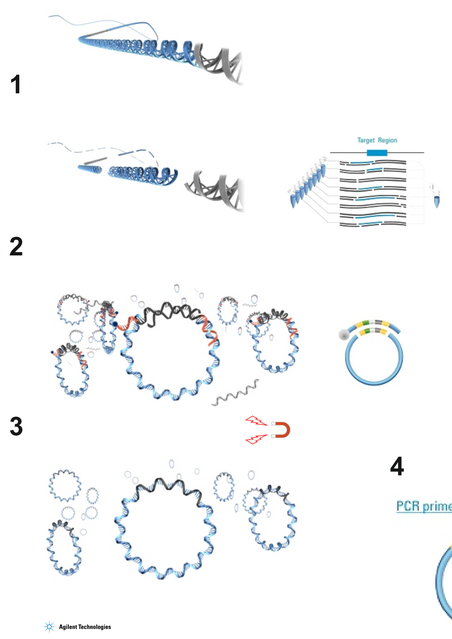
Ataxia: SCA NGS HP

- 140 ataxia genes (including pure ataxia, mitochondrial ataxia & metabolic ataxia genes)
- autosomal dominant & recessive, X -linked genes
- 582 kbp target region size
- ~ 2668 exons
- 25 496 fragments
- 1 403 kbp total size to sequence per sample
- ~ 22 samples per pool with 300x coverage

Paraplegia: HSP NGS HP

- 98 paraplegia genes (including all known SPG genes, spastic ataxia & other potential candidate genes)
- autosomal dominant & recessive, X -linked genes
- 267 kbp target region size
- ~ 1514 exons
- 13 250 fragments
- 764 kbp total size to sequence per sample
- ~ 39 samples per pool with 300x coverage

4 HaloPlex technology



Target enrichment with selector probes

1. Digestion of genomic DNA (defined cuts by restriction enzymes)
2. **Hybridisation of probes** (MIDs & seqeuncing motifs are incorporated)
3. Capture (purification) & Ligation
4. Amplification of circularised fragments

→ Up to 96 samples can be pooled and sequenced at once

5 Next -Generation -Sequencing device

Illumina MiSeq

Specifications

Medium throuput NGS system

- 2 x 150bp read length PE

→ covers average human exon length

- ~ 24h run time
- 25 - 34 Mio PE reads / run
- 8 - 15 Gbp output
- >80% of bases > Q30

6 Results

Sequencing performance

SCA NGS HP

> 28 ataxia samples processed

- 1,222 +/- 0,454 Mio total reads
- 91 +/- 2 % of bases > Q30
- 89 +/- 1 % of bases on target
- 189 +/- 32 bases average depth
- 96 +/- 1,6 % of bases ≥20 x covered

HSP NGS HP

> 9 paraplegia samples processed

- 1,338 +/- 0,387 Mio total reads
- 93 +/- 2 % of bases > Q30
- 84 +/- 1 % of bases on target
- 322 +/- 107 bases average depth
- 98 +/- 0,6 % of bases ≥20 x covered

Sequencing output

To get variants, our in-house bioinformatic pipeline mapped sequenced reads to the human genome (hg19) and annotated results to different databases using ANNOVAR.

Raw variant lists were then filtered for rare variants (in-house NGS database, 1000g, esp6500) and for functional relevance (non-synonymous, splicing).

SCA NGS HP

- 350 - 416 total variants per sample
- 14 - 26 rare variants per sample
- 9 rare variants validated by Sanger Sequencing
  - 4 likely pathogenic in 3 samples
    - 1x *CACNA1A* (SCA6) mutation (het)
    - 1x *FGF14* (SCA27) mutation (het)
    - 2x *SPG7* mutation (het)
- 10 samples without any relevant variant → WES analysis
- 15 samples with unclear variants (VUS3) → Validation

HSP NGS HP

- 156 - 169 total variants per sample
- 8 - 13 rare variants per sample
- 4 rare variants validated by Sanger Sequencing
  - 4 likely pathogenic in 3 samples
    - 1x *KIF1C* mutation (homo)
    - 3x *SPG11* mutation (1x homo / 2x het)
- 2 samples without any relevant variant → WES analysis
- 3 samples with unclear variants (VUS3) → Validation

7 Conclusion

HaloPlex target enrichment + MiSeq NGS sequencing

- is fast → 20 hours for library preparation (over night)
  → 28 hours for sequencing
- is robust → 96% / 98% of bases ≥20 x covered
  → 91% / 93% of bases > Q30
- is efficient to detect putative disease causing mutations in known and potential candidate genes
- is a recommended tool prior to WES in research projects
- is validated as a powerfull diagnostic tool

Acknowledgement

The research leading to these results has received funding from the European community's Seventh Framenwork Programme (FP7/2007-2013) under grant agreement n°2012-305121 „Integrated European -omics research prject for diagnosis and therapy in rare neuromuscular and neurodegenerative diseases (NEUROMICS)".

Correspondence

University Hospital of Tübingen
Institute of Medical Genetics and Applied Genomics
Calwerstrasse 7
D 72076 Tübingen / Germany
martin.schulze@med.uni-tuebingen.de
florian.harmuth@med.uni-tuebingen.de
peter.bauer@med.uni-tuebingen.de