Clinical Sequencing

Kevin Shianna, PhD
Senior Vice President, Sequencing Operations
New York Genome Center
Background

• Next Generation Sequencing
  – $1000 exome/$4000 genome
• Sequence a genome in a day (maybe hours?)
• Can we identify, annotate and interpret unique (rare) variants?
NGS to identify mutations in:

- Mendelian disease genes
- Undiagnosed genetic conditions
- Complex diseases
- Non-invasive prenatal testing (NIPT)
Targeted capture and massively parallel sequencing of 12 human exomes

Sarah B. Ng¹, Emily H. Turner¹, Peggy D. Robertson¹, Steven D. Flygare¹, Abigail W. Bigham², Choli Lee¹, Tristan Shaffer¹, Michelle Wong¹, Arindam Bhattacharjee⁴, Evan E. Eichler¹,³, Michael Bamshad², Deborah A. Nickerson¹ & Jay Shendure¹

-Freeman-Sheldon syndrome
-Autosomal dominant – MYH3
-June 2009

<table>
<thead>
<tr>
<th>Non-synonymous cSNP, splice site variant or coding indel (NS/SS/I)</th>
<th>FSS24895</th>
<th>FSS10208</th>
<th>FSS10066</th>
<th>FSS22194</th>
<th>Any 3 of 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>NS/SS/I not in dbSNP</td>
<td>4,510</td>
<td>3,284</td>
<td>2,765</td>
<td>2,479</td>
<td>3,768</td>
</tr>
<tr>
<td>NS/SS/I not in eight HapMap exomes</td>
<td>513</td>
<td>128</td>
<td>71</td>
<td>53</td>
<td>119</td>
</tr>
<tr>
<td>NS/SS/I neither in dbSNP nor eight HapMap exomes</td>
<td>799</td>
<td>168</td>
<td>53</td>
<td>21</td>
<td>160</td>
</tr>
<tr>
<td>...And predicted to be damaging</td>
<td>360</td>
<td>38</td>
<td>8</td>
<td>1 (MYH3)</td>
<td>22</td>
</tr>
<tr>
<td>...And predicted to be damaging</td>
<td>160</td>
<td>10</td>
<td>2</td>
<td>1 (MYH3)</td>
<td>3</td>
</tr>
</tbody>
</table>
Exome sequencing identifies *MLL2* mutations as a cause of Kabuki syndrome

Sarah B Ng¹,⁷, Abigail W Bigham²,⁷, Kati J Buckingham², Mark C Hannibal²,³, Margaret J McMillin², Heidi I Gildersleeve², Anita E Beck²,³, Holly K Tabor²,³, Gregory M Cooper¹, Heather C Mefford², Choli Lee¹, Emily H Turner¹, Joshua D Smith¹, Mark J Rieder¹, Koh-ichiro Yoshiura⁴, Naomichi Matsumoto⁵, Tohru Ohta⁶, Norio Niikawa⁶, Deborah A Nickerson¹, Michael J Bamshad¹–³ & Jay Shendure¹

- Kabuki Syndrome
- 10 probands
- Autosomal Dominant?
- 1/32,000
- April 2010
When this was applied to the exome data as a post hoc phenotype-ranked cases with a loss-of-function filter, in each of the four highest-ranked cases. After sequential analysis of lighted distinct, previously unidentified nonsense variants in splice-site disruption and frameshift compared to in-frame indel) zygotic twins with Kabuki syndrome. We then categorized the functional impact of the variants present would facilitate the prioritization of candidate genes identified by dbSNP129 or 1000 Genomes and control exomes (not in control exomes) or both (not in either); control exomes refer to those from 8 Hapmap3, 4 FSS3, 4 Miller2 and 10 EGP samples. The number of genes found using the union of the intersection of x individuals is given.

### Table 1 Number of genes common to any subset of x affected individuals.

<table>
<thead>
<tr>
<th>Subset analysis (any x of 10)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>NS/SS/I</td>
<td>12,042</td>
<td>8,722</td>
<td>7,084</td>
<td>6,049</td>
<td>5,289</td>
<td>4,581</td>
<td>3,940</td>
<td>3,244</td>
<td>2,486</td>
<td>1,459</td>
</tr>
<tr>
<td>Not in dbSNP129 or 1000 Genomes</td>
<td>7,419</td>
<td>2,697</td>
<td>1,057</td>
<td>488</td>
<td>288</td>
<td>192</td>
<td>128</td>
<td>88</td>
<td>60</td>
<td>34</td>
</tr>
<tr>
<td>Not in control exomes</td>
<td>7,827</td>
<td>2,865</td>
<td>1,025</td>
<td>399</td>
<td>184</td>
<td>90</td>
<td>50</td>
<td>22</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>Not in either</td>
<td>6,935</td>
<td>2,227</td>
<td>701</td>
<td>242</td>
<td>104</td>
<td>44</td>
<td>16</td>
<td>6</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Is loss-of-function (nonsense or frameshift indel)</td>
<td>753</td>
<td>49</td>
<td>7</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

The number of genes with at least one nonsynonymous variant (NS), splice-site acceptor or donor variants (SS) or coding indel (I) are listed under various filters. Variants were filtered by presence in dbSNP or 1000 Genomes (not in dbSNP129 or 1000 genomes) and control exomes (not in control exomes) or both (not in either); control exomes refer to those from 8 Hapmap3, 4 FSS3, 4 Miller2 and 10 EGP samples. The number of genes found using the union of the intersection of x individuals is given.

### Table 2 Number of genes common in sequential analysis of phenotypically ranked individuals

<table>
<thead>
<tr>
<th>Sequential analysis</th>
<th>1</th>
<th>+2</th>
<th>+3</th>
<th>+4</th>
<th>+5</th>
<th>+6</th>
<th>+7</th>
<th>+8</th>
<th>+9</th>
<th>+10</th>
</tr>
</thead>
<tbody>
<tr>
<td>NS/SS/I</td>
<td>5,282</td>
<td>3,850</td>
<td>3,250</td>
<td>2,354</td>
<td>2,028</td>
<td>1,899</td>
<td>1,772</td>
<td>1,686</td>
<td>1,600</td>
<td>1,459</td>
</tr>
<tr>
<td>Not in dbSNP129 or 1000 Genomes</td>
<td>687</td>
<td>214</td>
<td>145</td>
<td>84</td>
<td>63</td>
<td>54</td>
<td>42</td>
<td>40</td>
<td>39</td>
<td>34</td>
</tr>
<tr>
<td>Not in control exomes</td>
<td>675</td>
<td>134</td>
<td>50</td>
<td>26</td>
<td>13</td>
<td>13</td>
<td>8</td>
<td>5</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Not in either</td>
<td>467</td>
<td>89</td>
<td>34</td>
<td>18</td>
<td>9</td>
<td>8</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Is loss-of-function (nonsense/frameshift indel)</td>
<td>25</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Variants were filtered as in Table 1. Exomes were added sequentially to the analysis by ranked phenotype; for example, column "+3" shows the number of genes at the intersection of the three top ranked cases (Supplementary Fig. 1). The gene with at least one NS/SS/I in all individuals is MUC16, which is very likely to be a false positive due to its extreme length (14,507 amino acids).
Clinical sequencing: best case scenario
Aug 2010

Making a definitive diagnosis: Successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease

Elizabeth A. Worthey, PhD1,2, Alan N. Mayer, MD, PhD2,3, Grant D. Syverson, MD2, Daniel Helbling, BSc1, Benedetta B. Bonacci, MSc2, Brennan Decker, BSc1, Jaime M. Serpe, BSc2, Trivikram Dasu, PhD2, Michael R. Tschannen, BSc1, Regan L. Veith, MSc2, Monica J. Basehore, PhD4, Ulrich Broeckel, MD, PhD1,2,3, Aoy Tomita-Mitchell, PhD1,2,3, Marjorie J. Arca, MD3,5, James T. Casper, MD2,3, David A. Margolis, MD2,3, David P. Bick, MD1,2,3, Martin J. Hessner, PhD1,2, John M. Routes, MD2,3, James W. Verbsky, MD, PhD2,3, Howard J. Jacob, PhD1,2,3,6, and David P. Dimmock, MD1,2,3

Nicholas Volker
Clinical sequencing: best case scenario

- 15 month old presented with Crohn disease like illness
- Age/severity suggested immune defect

- At 4 years of age
  - Hundreds of surgeries
  - Over 700 days in the hospital
  - Tested multiple candidate genes = normal
  - Treatments working, but not long term solution

- Immune reconstitution?
  - Aggressive approach
  - Unknown mechanism
Clinical sequencing: best case scenario

- Exome sequencing
  - Identified mutation in XIAP gene
  - X-linked inhibitor of apoptosis
  - Mutations known to cause XLP2
  - X-linked lymphoproliferative disease
  - Not previously associated with colitis

Table 1, category A, provides a summary of the total numbers and numbers of novel variations broken down by both location and variation class (insertions, deletions, and substitutions). As expected, the majority of variants identified were substitutions; insertions were the least common across all categories. A larger percentage of the novel variants were insertions or deletions rather than substitutions when compared with the previously identified variants. A small number of these nonsynonymous substitutions resulted in the production of stop codons; all homozygous examples were either previously identified or resulted in a stop in a protein commonly found in the population to harbor stop codons (Table 1, category A).

To reduce the search space, we hypothesized, based on the severity and unique clinical presentation, that this case was likely a recessive disorder caused by a hemizygous or homozygous mutation, or compound heterozygote. Sixty-six genes containing potential compound heterozygous mutations (where variants were nonsynonymous and predicted to be damaging by Polyphen, a tool that predicts the impact of an amino acid substitution on the structure and function of a human protein) were identified (Table 1, category B) and investigated; none of these candidates remained after exclusion based on novelty and sequence conservation. Seventy homozygous/hemizygous non-synonymous variants were identified (Table 1C); eight were novel (when compared against all publicly available data sets) and predicted to be damaging by PolyPhen. Only two of these were highly conserved. One variant, in GSTM1, was excluded because this gene has a high null genotype frequency in the population.
Clinical sequencing: best case scenario

- Functionality
  - Central role in proinflammatory response
  - Activation of NFkB followed by activation of proinflammatory cytokines
  - NOD signaling/mediated programmed cell death

- Treatment
  - Hemopoietic progenitor cell transplant
    - 42 days posttransplant – eat and drank normally
  - Complete resolution of colitis
  - No recurrence of gastrointestinal disease
Filtering process

- Quality scores/coverage
- dbSNP
- 1000 Genomes
- Exome variant server (EVS)
  - Over 5000 exomes (by frequency)
- Database of genomic variants (DGV)
- Affected/unaffected/trios
- Additional probands/families

- Online Medelian Inheritance in Man (OMIM)
- Human Gene Mutation Database (HGMD)
- PolyPhen-2/SIFT
Whole-Genome Sequencing for Optimized Patient Management

Matthew N. Bainbridge,¹,² Wojciech Wiszniewski,³ David R. Murdock,¹ Jennifer Friedman,⁴,⁵ Claudia Gonzaga-Jauregui,³ Irene Newsham,¹ Jeffrey G. Reid,¹ John K. Fink,⁶,⁷ Margaret B. Morgan,¹ Marie-Claude Gingras,¹ Donna M. Muzny,¹ Linh D. Hoang,⁸ Shahed Yousaf,⁸ James R. Lupski,¹,³,⁹,10 Richard A. Gibbs¹,³*  

Whole-genome sequencing of patient DNA can facilitate diagnosis of a disease, but its potential for guiding treatment has been under-realized. We interrogated the complete genome sequences of a 14-year-old fraternal twin pair diagnosed with dopa (3,4-dihydroxyphenylalanine)–responsive dystonia (DRD; Mendelian Inheritance in Man #128230). DRD is a genetically heterogeneous and clinically complex movement disorder that is usually treated with L-dopa, a precursor of the neurotransmitter dopamine. Whole-genome sequencing identified compound heterozygous mutations in the SPR gene encoding sepiapterin reductase. Disruption of SPR causes a decrease in tetrahydrobiopterin, a cofactor required for the hydroxylase enzymes that synthesize the neurotransmitters dopamine and serotonin. Supplementation of L-dopa therapy with 5-hydroxytryptophan, a serotonin precursor, resulted in clinical improvements in both twins.

-Fraternal twins (male/female) – 14 years old
-Dopa responsive dystonia
-Genetically heterogeneous
-Clinically complex movement disorder
-L-Dopa treatment
BH4 cofactor synthesis, leading to disruption in the production of the neurotransmitters dopamine, noradrenaline, and serotonin. In addition to DRD in the probands, the family has a history of depression and fibromyalgia on either side of the pedigree. Segregation of the two SPR mutations is shown for all individuals evaluated.

**Fig. 2.** Pedigree of a family segregating recessive DRD, depression, and fibromyalgia. Pedigree of the family with the two DRD-affected probands (shaded), male and female fraternal twins. Their DRD is due to disruption of SPR activity resulting in impaired
Dopa Responsive Dystonia (DRD)

<table>
<thead>
<tr>
<th>Nucleotide variants</th>
<th>IV-2</th>
<th>IV-3</th>
<th>Shared</th>
</tr>
</thead>
<tbody>
<tr>
<td>All variants</td>
<td>2,427,038</td>
<td>2,504,162</td>
<td>1,631,770</td>
</tr>
<tr>
<td>% dbSNP129</td>
<td>88.7</td>
<td>88.1</td>
<td></td>
</tr>
<tr>
<td>Variant density (bp⁻¹)</td>
<td>1/1112</td>
<td>1/1078</td>
<td></td>
</tr>
<tr>
<td>Coding</td>
<td>13,352</td>
<td>14,961</td>
<td>9531</td>
</tr>
<tr>
<td>Nonsynonymous</td>
<td>6432</td>
<td>7141</td>
<td>4605</td>
</tr>
<tr>
<td>Rare nonsynonymous</td>
<td>174</td>
<td>175</td>
<td>77</td>
</tr>
<tr>
<td>Candidate genes</td>
<td>6</td>
<td>9</td>
<td>3</td>
</tr>
</tbody>
</table>
Dopa Responsive Dystonia (DRD)

- Compound het in SPR
- Supplement with 5-HTP (L-dopa)

Guanosine triphosphate (GTP)

\[ \text{GCH1} \]

Dihydroneopterin triphosphate

\[ \text{PTPS} \]

6-Pyruvoyl-tetrahydropterin

\[ \text{SPR} \]

Tetrahydrobiopterin (BH4)

\[ \text{TH} \]

L-dopa

Tyrosine

TPH

Tryptophan

PAH

Phenylalanine

Tyrosine

5-Hydroxytryptophan (5-HTP)

Serotonin

5-Hydroxyindoleacetic acid

Dopamine

Homovanillic acid

Norepinephrine

5-Hydroxytryptophan

Quinoid-dihydrobiopterin

DHPR

PCD

Pterin-4a-carbinolamine

Guanosine triphosphate (GTP)

Dihydroneopterin triphosphate

6-Pyruvoyl-tetrahydropterin

Tetrahydrobiopterin (BH4)

L-dopa

Tyrosine

TPH

Tryptophan

PAH

Phenylalanine

Tyrosine

5-Hydroxytryptophan (5-HTP)

Serotonin

5-Hydroxyindoleacetic acid

Dopamine

Homovanillic acid

Norepinephrine

5-Hydroxytryptophan

Quinoid-dihydrobiopterin

DHPR

PCD

Pterin-4a-carbinolamine
Trio sequencing

• 40 or so de novo coding variants per genome
  – After close inspection/validation only a fraction are real (most only have 0-3)
• Rare diseases and Complex diseases
De novo mutations in ATP1A3 cause alternating hemiplegia of childhood

Alternating Hemiplegia of Childhood (AHC)

- Intermittent episodes of hemiplegia on alternating sides of the body, abnormal eye movement
- Symptoms before 18 months of age.
- Developmental delay
- 1 in 1,000,000
Exome sequencing of seven AHC trios

• De novo mutations in ATP1A3
  – Sodium potassium ATPase subunit
• All nonsynonymous mutations
• Mutations in ATP1A3 previously known to cause rapid-onset dystonia-parkinsonism (DYT12)
  – Loss/decreased protein levels
• AHC-causing mutations
  – Consistent reduction of ATPase activity
  – Normal protein expression level
Two AHC causing mutations affect amino acids also affected by DYT12-causing mutations
Clinical application of exome sequencing in undiagnosed genetic conditions

Anna C Need, Vandana Shashi, Yuki Hitomi, Kelly Schoch, Kevin V Shianna, Marie T McDonald, Miriam H Meisler, David B Goldstein

INTRODUCTION

There is considerable interest in the use of next-generation sequencing (NGS) to provide genetic diagnoses of conditions for which no genetic cause is currently recognised. This study provides evidence that NGS can have high success rates in a clinical setting that includes patients with unexplained and apparent genetic conditions, along with congenital anomalies and/or intellectual disabilities. This can be representative of a clinical sample of undiagnosed genetic conditions, in that they were not predetermined criteria were met. Importantly, the patients were chosen to represent the clinical setting in which NGS is most likely to be used, and they were not selected for genetic tractability or phenotypic features, but rather enrolled any undiagnosed genetic conditions along with congenital anomalies and/or intellectual disabilities.

METHODS

Subjects were not eligible if the mother was pregnant, if they had been normal; and (3) no evidence of effects of any causative genetic factors. Subjects were enrolled in Center for Human Genome Variation following confirmation of detected variants in genes known to cause Mendelian disease.

RESULTS

There are 12 families (child, mother and father) enrolled in the study. The research protocol was approved by the Duke Institutional Review Board, and all human participants provided informed consent.

CONCLUSIONS

Clinical application of exome sequencing in undiagnosed genetic conditions can have high success rates in a clinical setting, but also highlights key challenges.
Table 1  Demographic and clinical features of sequenced patients

<table>
<thead>
<tr>
<th>Trio</th>
<th>Sex</th>
<th>Age</th>
<th>Race</th>
<th>Symptoms</th>
<th>Genetic tests performed clinically before enrolment in study</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>M</td>
<td>8</td>
<td>Indian</td>
<td>Developmental delay, possible autism, microcephaly, dysmorphic features, spine abnormalities, sensorineural hearing loss</td>
<td>Chromosome microarray (paternally inherited 15q13.3 dup), Fragile X</td>
</tr>
<tr>
<td>2</td>
<td>M</td>
<td>3</td>
<td>European-American</td>
<td>Developmental delay, multifocal epilepsy, involuntary movements, abnormal liver function, absent tears</td>
<td>Chromosomes, chromosome microarray, Niemann-Pick type C, hepatocerebral mDNA depletion panel (POLG1, DGU0K, MPV17), ataxia with oculomotor apraxia type 2 (SETX), Allgrove Syndrome, ataxia telangetasia (ATM), Rett (MECP2), alpha—1 antitrypsin (AAT), congenital disorder of glycosylation (transferrin isoelectric focusing and N-glycan analysis), metabolic tests (Tay Sachs, Sandhoff, mannosidosis, mucolipidosis II, Krabbe, metachromatic leukodystrophy, adrenoleukodystrophy, GAMT, plasma amino acids, plasma acylcarnitine, urine organic acids).</td>
</tr>
<tr>
<td>3</td>
<td>M</td>
<td>3</td>
<td>European-American</td>
<td>Developmental delay, autism, coarctation of the aorta, tethered cord, congenital nystagmus and strabismus</td>
<td>Chromosome microarray (maternally inherited 15q26.3 deletion), Smith-Lemli-Opitz, Aarskog</td>
</tr>
<tr>
<td>4</td>
<td>F</td>
<td>adult</td>
<td>European-American</td>
<td>multiple congenital abnormalities and macular degeneration</td>
<td>Chromosome microarray (2 stretches of loss of heterozygosity on chromosome 2), Fragile X (premutation carrier).</td>
</tr>
<tr>
<td>5</td>
<td>F</td>
<td>12</td>
<td>European-American</td>
<td>Severe intellectual disability, autism, bilateral hyperpronated feet, facial dysmorphisms</td>
<td>Chromosomes, chromosome microarray, Rett, Angelman methylation, Fragile X, Cohen Syndrome</td>
</tr>
<tr>
<td>6</td>
<td>M</td>
<td>18</td>
<td>European-American</td>
<td>Intellectual disability, epilepsy, panhypopituitarism, hypertension, bifid great toe, vertebral segmentation anomalies and sagittal cleft of the vertebra, hypoplastic 13th rib, and delayed bone age</td>
<td>Chromosomes, chromosome microarray, Borgeson-Foorsman-Lehman syndrome</td>
</tr>
<tr>
<td>7</td>
<td>M</td>
<td>2</td>
<td>European-American</td>
<td>Microcephaly, facial asymmetry, acyanotic Tetralogy of Fallot; history of small muscular ventricular septal defect; right aortic arch with mirror image branching; malformed right ear with hearing loss, bifid uvula, cleft soft palate</td>
<td>Chromosome microarray, CHARGE (CHD7)</td>
</tr>
<tr>
<td>8</td>
<td>M</td>
<td>16</td>
<td>European-American</td>
<td>Severe intellectual disability, dysmorphic features evident, bicuspid aortic valve, bilateral coronal craniosynostoses, quadriplegic cerebral palsy, bilateral inguinal hernias, G-tube placement and obstructive sleep apnoea</td>
<td>Chromosome microarray, craniosynostosis syndromes (FGFR2), non-syndromic craniosynostosis (FGFR3) Saethre-Chotzen syndrome (TWIST)</td>
</tr>
<tr>
<td>9</td>
<td>F</td>
<td>4</td>
<td>Algerian</td>
<td>Developmental delay, bilateral congenital cataracts and strabismus, ventricular and atrial septal defects, a unilateral clubfoot, and unilateral choanal atresia</td>
<td>Chromosome microarray (Long stretch of loss of heterozygosity on chromosome 17), CHARGE (CHD7), PAX6, 7-dehydrocholesterol and cholesterol levels</td>
</tr>
<tr>
<td>10</td>
<td>M</td>
<td>11</td>
<td>European-American</td>
<td>Attention deficit hyperactivity disorder, language delays, coarse facial features, bilateral mandibular cysts, low muscle tone</td>
<td>Chromosome microarray, Costello (H-RAS), Gorlin (PTCH), Comprehensive Noonan sequencing array (BRAF, HRAS, Kras MAPT2K1, MAPT2K2, PTPN11, RAF1, SHOC2 and SOS1), MPS panel</td>
</tr>
<tr>
<td>11</td>
<td>M</td>
<td>9</td>
<td>European-American</td>
<td>Severe intellectual disability, developmental delay, seizures/infantile spasms, hypotonia and minor dysmophisms</td>
<td>Chromosomes, chromosome microarray (familial Xp11.4 duplication), acylcarnitine profile, plasma amino acids, urine organic acids, creatine/creatinine analysis in urine and blood</td>
</tr>
<tr>
<td>12</td>
<td>F</td>
<td>4</td>
<td>European-American</td>
<td>Speech delay, borderline microcephaly, failure to thrive, dysplastic nails, ventricular septal defect and hip dysplasia</td>
<td>Chromosomes, chromosome microarray</td>
</tr>
</tbody>
</table>
Trio/Patient 5

- De novo mutation in TCF4
  - Transcription Factor 4
  - Pitt Hopkins syndrome (PHS)
  - Haploinsufficiency of TCF4
  - Atypical PHS?

![Figure 2](image_url)  
*Figure 2* Expression of TCF4 variant and wild-type (WT) protein in COS-7 cells. The variant protein (V) is only seen in the presence of proteasome inhibitors. GAPDH, glyceraldehyde 3-phosphate dehydrogenase.
Important lessons

• Variability in clinical presentation of genetic disorders
• Ability to interpret unique variants is a work in progress
• Functional work is necessary
Genome Sequencing

a) Sequencing affected individuals in families

Sequencing affected relatives
Identify shared variants that are likely to cause disease

Genotype candidate variants
Look for co-segregation with disease in family

b) Extreme-trait sequencing

Whole-genome sequencing
Rare variant enriched in population with extreme phenotype

Targeted genotype
Rare variant associates with phenotype in larger population

Variant of large effect at low frequency in general population

Cirulli ET& Goldstein DB. Nature Review Genetics, 2010
LETTER

Patterns and rates of exonic de novo mutations in autism spectrum disorders

Benjamin M. Neale1,2, Yan Kou3,4, Li Liu5, Avi Ma'ayan3, Kaitlin E. Samocha3,2, Aniko Sabo6, Chiao-Feng Lin7, Christine Stevens3, Li-San Wang7, Vladimir Makarov4,8, Paz Polak2,7, Seungtaek Yoon1,8, Jared Mauger2, Emily L. Crawford18, Nicholas G. Campbell19, Evan T. Geller1, Otto Valladares2, Chad Schafer2, Han Liu1, Tu Zhao1, Guokun Cai1,8, Jiyue Liu1,8, Ruth Dannenfelser2, Omar Jabado15, Zuleyma Peralta15, Uma Nagaswamy2, Donna Muzny4, Jeffrey G. Reid9, Irene Newsham9, Yuanqiong Wu6, Lora Lewis6, Yi Han6, Benjamin F. Voight12,33, Elaine Lim1,2, Elizabeth Rossin1,2, Andrew Kirby1,2, Jason Flannick2, Menachem Fromer1,2, Khalid Shakir2, Tim Fennell1, Kiran Garimella2, Eric Banks2, Ryan Poppin2, Stacey Gabriel2, Mark DePristo7, Jack R. Wimbish1,4, Braden E. Boone1,4, Shawn E. Levy14, Catalina Betancur15, Shamil Sunyaev1,2,9, Eric Boerwinkle6,16, Joseph D. Buxbaum1,3,32,17, Edwin H. Cook Jr18, Bernie Devlin19, Richard A. Gibbs6, Kathryn Roeder2, Gerard D. Schellenberg2, James S. Sutcliffe10 & Mark J. Daly1,2

LETTER

De novo mutations revealed by whole-exome sequencing are strongly associated with autism

Stephan J. Sanders1, Michael T. Murtha1, Abha R. Gupta24, John D. Murdock1,6, Melanie J. Raubeson1,6, A. Jeremy Willsey1,6, A. Gulhan Ercan-Sencicek6, Nicholas M. DiLullo2,6, Neel Grover2,6, Jason A. Stein2, Michael F. Walker1, Gordon T. Ober1, Nicole A. Teran1, Youuen Song1, Paul E. Fishaw2, Ryan C. Murtha1, Murat Chof2, John D. Overton1, Robert D. Bjornson3, Nicholas J. Carriero1, Kyle A. Meyer1, Kaya Bilgunar2, shriram M. Mane3, Narendra Sestian2, Richard P. Lifton3, Murat Gunes1, Kathryn Roeder2, Daniel H. Geschwind2, Bernie Devlin19 & Matthew W. State1

LETTER

Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations

Brian J. O’Roak1, Laura Vives1, Santosh G. Girirajan1, Emre Karakoc1, Niklas Krumm1, Bradley P. Coe1, Roie Levy1, Arthur Ko1, Cholo Lee1, Joshua D. Smith2, Emily H. Turner1, Ian B. Stanaway1, Benjamin Vernot3, Maika Malig5, Carl Baker2, Shri R. Roy5, Joshua M. Akey1, Elhanan Borenstein1,2,3,4, Mark J. Rieder1, Deborah A. Nickerson1, Raphael Bernier2, Jay Shendure1 & Evan E. Eichler1,5
We observed recurrent, protein-terminating mutations identified here are missense (p.Tyr23Cys and p.Thr135Ile) at highly conserved positions predicted to disrupt protein function, although there is evidence of mosaicism for the former mutation in CHD8 (ref. 16). ASD has been found in as many as two-thirds of children with CHARGE, indicating that CHD8 (OMIM 214800), a known binding partner of CHD7 (OMIM 605309) and NTNG1 (netrin G1) and RPS6KA3 (RPS6 kinase 3), and microcephalic, consistent with previous studies of dominant 1, OMIM 156200), and missense variants in loci associated with intellectual disability genes (mental retardation, autosomal dominant 1, OMIM 303600) and macrocephaly/autism (17q12 deletion syndrome) and MBD5 (methyl-CpG binding protein 5), indicating diverse phenotypic outcomes at this locus depending on mutation mechanism.

Among the de novo mutations identified here are missense mutations at SETBP1 (18q12.3) in two unrelated cases and paternal estimated age at conception versus the number of observed loci-specific mutation rates, the probability of identifying two independent mutations in our sample set is low (uncorrected, P = 0.01) with increased number of events occurring on the paternally inherited haplotypes.

Covariant analysis of the samples with CNV data showed that this finding was strengthened, but not exclusively driven, by the presence of either Mendelian or ASD loci (Supplementary Discussion). From our full list of genes among the 62 top ASD risk containing CNVs that intersect genes mapping to the Down's syndrome critical region (Supplementary Figure 7, Table 12), or CNVs intersecting with Mendelian disease loci (Supplementary Table 3) as well as inherited hemizygous mutations of clinical significance (Supplementary Table 9).
locus heterogeneity

Number of genes with a mutation in two or more individuals

Number of cases

- 50 genes
- 500 genes
- 1000 genes
- 5000 genes
Clinical Sequencing in the NICU
Rapid Whole-Genome Sequencing for Genetic Disease Diagnosis in Neonatal Intensive Care Units

Carol Jean Saunders,1,2,3,4,5* Neil Andrew Miller,1,2,4* Sarah Elizabeth Soden,1,2,4* Darrell Lee Dinwiddie,1,2,3,4,5* Aaron Noll,1 Noor Abu Alnadi,4 Nevene Andraws,3 Melanie LeAnn Patterson,1,3 Lisa Ann Krivohlavek,1,3 Joel Fellis,6 Sean Humphray,6 Peter Saffrey,6 Zoya Kingsbury,6 Jacqueline Claire Weir,6 Jason Betley,6 Russell James Grocock,6 Elliott Harrison Margulies,6 Emily Gwendolyn Farrow,1 Michael Artman,2,4 Nicole Pauline Safina,1,4 Joshua Erin Petrik,2,3 Kevin Peter Hall,6 Stephen Francis Kingsmore1,2,3,4,5†

Monogenic diseases are frequent causes of neonatal morbidity and mortality, and disease presentations are often undifferentiated at birth. More than 3500 monogenic diseases have been characterized, but clinical testing is available for only some of them and many feature clinical and genetic heterogeneity. Hence, an immense unmet need exists for improved molecular diagnosis in infants. Because disease progression is extremely rapid, albeit heterogeneous, in newborns, molecular diagnoses must occur quickly to be relevant for clinical decision-making. We describe 50-hour differential diagnosis of genetic disorders by whole-genome sequencing (WGS) that features automated bioinformatic analysis and is intended to be a prototype for use in neonatal intensive care units. Retrospective 50-hour WGS identified known molecular diagnoses in two children. Prospective WGS disclosed potential molecular diagnosis of a severe GJB2-related skin disease in one neonate; BRAT1-related lethal neonatal rigidity and multifocal seizure syndrome in another infant; identified BCL9L as a novel, recessive visceral heterotaxy gene (HTX6) in a pedigree; and ruled out known candidate genes in one infant. Sequencing of parents or affected siblings expedited the identification of disease genes in prospective cases. Thus, rapid WGS can potentially broaden and foreshorten differential diagnosis, resulting in fewer empirical treatments and faster progression to genetic and prognostic counseling.

- Hard to diagnose
- Time sensitive!
- Symptom and sign assisted genome analysis (SSAGA)
50 hrs from patient consent to diagnosis

- Obtain consent and blood sample
- Prepare sequencing library
  Enter clinical findings into SSAGA
- HiSeq 2500 2 x 100 bp sequencing
- CASAVA base calling
  RUNES variant annotation
- SSAGA-delimited variant analysis and interpretation
- Verbal interim report of diagnosis pending CLIA confirmation
Non-Invasive Prenatal Testing (NIPT)
Noninvasive prenatal diagnosis of fetal chromosomal aneuploidy by massively parallel genomic sequencing of DNA in maternal plasma


© 2008 by The National Academy of Sciences of the USA
plasma, the proportional contribution of plasma DNA sequences maternal and the fetal genomes are evenly represented in maternal DNA fragments in maternal plasma. Furthermore, if both the sequence DNA fragments originating from each chromosome. When these assumptions hold, then the % chrN values should be generated sequence reads for the small fraction of fetal DNA in maternal plasma alongside the background maternal DNA. Second, there should be no major bias in the ability to generate a similar interchromosomal distribution to that in the original maternal plasma of euploid pregnancies. Third, there should be no major bias in the ability to pool of plasma DNA fragments captured for sequencing from an aneuploid chromosome could be more readily achieved. chromosomal aneuploidy detection, a number of assumptions need from the mean and standard deviation of % chr21 values obtained from maternal plasma, the number of quantifiable sequences would be much greater than the number of DNA molecules that could serve the amount of that chromosome. Therefore, for any fixed volume method is used, potentially every DNA fragment originating from independent of any particular gene locus to quantify the amount of expected from the mean of a reference data set. Hence, for a T21 interest is chrN. To determine if a tested maternal plasma sample of the tested sample. The -score refers to the number of standard deviations from the mean of a reference data set. Hence, for a T21 poner-ligated DNA fragments both before and after (i.e., two rounds expected to be higher for pregnancies with an aneuploid fetus (cases E–H a large enough pool of plasma DNA sequences, we hypothesize that we would be able to discriminate perturbations in size of each chromosome in the human genome. If the % chrN per chromosome should in turn bear correlation with the relative of amplification) a gel electrophoresis-based size fractionation step identified through mapping to the human reference genome by bioinformatics alignment analysis for the Illumina Genome Analyzer, from an aneuploid pregnancy, we calculated the -score of % chr21 was expected when compared with -score for % chr21 was determined using the repeat-masked reference human genome with no mismatch, i.e., sequences that could be mapped to just one location in the databases (ELAND) software. The purpose of the alignment was used, for this purpose. This study, we demonstrate the use of the Solexa sequencing technique (Illumina) (18) for this purpose. We previously (15) proposed that the recently available massively parallel genomic sequencing (MPGS) platforms (16, 17) might be adaptable as an approach to quantify DNA sequences for the identification of unique sequences mapped to each chromosome by dividing the U0–1 sequences as U0–1 deemed as a unique sequence in the human genome. We termed this repeat-masked reference human genome with no mismatch, i.e., sequences in the data output files of the ELAND sequence alignment software (Illumina) (see supporting information (SI) text) sequences mapped to of unique sequences mapped to a chromosome.

\[ \text{% chrN} = \frac{\text{Unique count for chrN}}{\text{Total unique count}} \]

\[ \text{chrN z-score for test sample} = \frac{\% \text{chrN}_{\text{sample}} - \text{mean} \% \text{chrN}_{\text{reference}}}{\text{S.D.} \% \text{chrN}_{\text{reference}}} \]
Noninvasive Whole-Genome Sequencing of a Human Fetus

Jacob O. Kitzman,1* Matthew W. Snyder,1 Mario Ventura,1,2 Alexandra P. Lewis,1 Ruolan Qiu,1 LaVone E. Simmons,3 Hilary S. Gammill,3,4 Craig E. Rubens,5,6 Donna A. Santillan,7 Jeffrey C. Murray,8 Holly K. Tabor,5,9 Michael J. Bamshad,1,5 Evan E. Eichler,1,10 Jay Shendure1*

Analysis of cell-free fetal DNA in maternal plasma holds promise for the development of noninvasive prenatal genetic diagnostics. Previous studies have been restricted to detection of fetal trisomies, to specific paternally inherited mutations, or to genotyping common polymorphisms using material obtained invasively, for example, through chorionic villus sampling. Here, we combine genome sequencing of two parents, genome-wide maternal haplotyping, and deep sequencing of maternal plasma DNA to noninvasively determine the genome sequence of a human fetus at 18.5 weeks of gestation. Inheritance was predicted at $2.8 \times 10^6$ parental heterozygous sites with 98.1% accuracy. Furthermore, 39 of 44 de novo point mutations in the fetal genome were detected, albeit with limited specificity. Subsampling these data and analyzing a second family trio by the same approach indicate that parental haplotype blocks of ~300 kilo–base pairs combined with shallow sequencing of maternal plasma DNA is sufficient to substantially determine the inherited complement of a fetal genome. However, ultradeep sequencing of maternal plasma DNA is necessary for the practical detection of fetal de novo mutations genome-wide. Although technical and analytical challenges remain, we anticipate that noninvasive analysis of inherited variation and de novo mutations in fetal genomes will facilitate prenatal diagnosis of both recessive and dominant Mendelian disorders.
Table 1. Summary of sequencing. Individuals sequenced, type of starting material, and final fold coverage of the reference genome after discarding PCR or optical duplicate reads.

<table>
<thead>
<tr>
<th>Individual</th>
<th>Sample</th>
<th>Depth of coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mother (I1-M)</td>
<td>Plasma (5 ml, gestational age 18.5 weeks)</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td>Whole blood (&lt;1 ml)</td>
<td>32</td>
</tr>
<tr>
<td>Father (I1-P)</td>
<td>Saliva</td>
<td>39</td>
</tr>
<tr>
<td>Offspring (I1-C)</td>
<td>Cord blood at delivery</td>
<td>40</td>
</tr>
</tbody>
</table>

Figure A

- Plasma
- WGS + haplotypes
- WGS (validation)
- WGS
Maternal plasma DNA sequences were ~13% fetal-derived on the basis of assuming 13% fetal content). To asse
Assuming 13% fetal content, we performed a site-by-site log-odds test to assess these, taking the observation of one or more reads matching the paternal-specific label as evidence of its transmission and, conversely, the lack of such observations as evidence of nontransmission (Fig. 1C). In contrast to maternal-only heterozygous sites, this simple site-by-site model was sufficient to correctly predict inheritance at $1.1 \times 10^6$ paternal-only heterozygous sites with 96.8% accuracy (Table 2). We anticipate that accuracy could likely be improved by deeper sequence coverage of the maternal plasma DNA (fig. S2) or, alternatively, by taking a haplotype-based approach if high-molecular-weight genomic DNA from the father is available.

We next considered transmission at sites heterozygous in both parents. We predicted maternal transmission at such shared sites phased using neighboring maternal-only heterozygous sites in the same haplotype block. This yielded predictions at 576,242 of 631,721 (91.2%) of shared heterozygous sites with an estimated accuracy of 98.7% (Table 2). Although we did not predict paternal transmission at these sites, we anticipate that analogous to the case of maternal transmission, this could be done with high accuracy given paternal haplotypes. We note that shared heterozygous sites primarily correspond to common alleles (fig. S3), which are less likely to contribute to Mendelian disorders in nonconsanguineous populations.

---

### Figure 1. Experimental approach

- **A** Sequenced individuals in a family trio. Maternal plasma DNA sequences were ~13% fetal-derived on the basis of read depth at chromosome Y and alleles specific to each parent. WGS, whole-genome shotgun.
- **B** Inheritance of maternally heterozygous alleles inferred using long haplotype blocks. Among plasma DNA sequences, maternal-specific alleles are more abundant when transmitted (expected, 50% versus 43.5%), but there is substantial overlap between the distributions of allele frequencies when considering sites in isolation (left histogram: yellow, shared allele transmitted; green, maternal-specific allele transmitted). Taking average allele balances across haplotype blocks (right histogram) provides much greater separation, permitting more accurate inference of maternally transmitted alleles.
- **C** Histogram of fractional read depth among plasma data at paternal-specific heterozygous sites. In the overwhelming majority of cases when the allele specific to the father was not detected, the opposite allele had been transmitted (96.8%, $n = 561,552$).
- **D** De novo missense mutation in the gene ACMSD detected in 3 of 93 maternal plasma reads and later validated by PCR and resequencing. The mutation, which is not observed in dbSNP nor among coding exons sequenced from >4000 individuals as part of the National Heart, Lung, and Blood Institute Exome Sequencing Project (http://evs.gs.washington.edu), creates a leucine-to-proline substitution at a site conserved across all aligned mammalian genomes (University of California, Santa Cruz, Genome Browser) in a gene implicated in Parkinson's disease by genome-wide association studies (25).
NGS caveats for clinical sequencing

• Short read technology/aligning to reference genome

• SNV vs indels, CNV......
• Small repeats/repetitive sequence
• Phasing

• Interpretation of unique variants!