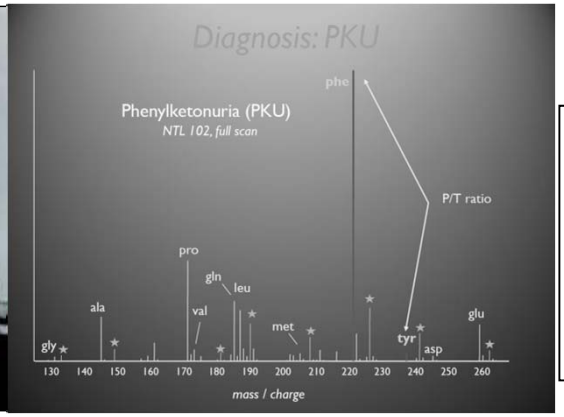

# Second-tier DNA Confirmation of Newborn Screening by Targeted Next Generation Sequencing

Edwin Naylor, Ph.D. MPH
Andy Bhattacharjee , Ph.D.
Erik Puffenberger, Ph.D.; Kevin Strauss, MD; Holmes Morton, MD

PARABASE GENOMICS

# Newborn Screening & Clinical Genomics

1961 — 1990's — 2010-2012




Robert Guthrie develops simple Newborn Screening (NBS)

Development of automated MS/MS screening across several disorders

Current *de facto* standard

SCIENCE VOL 330 17 DECEMBER 2010

BREAKTHROUGH OF THE YEAR | NEWSFOCUS

HUMAN GENOMICS

**Carrier Testing for Severe Childhood Recessive Diseases by Next-Generation Sequencing**

Callum J. Bell,[1]* Darrell L. Dinwiddie,[1,2]* Neil A. Miller,[1,2] Shannon L. Hateley,[1] Elena E. Ganusova,[1] Joann Mudge,[1] Ray J. Langley,[1] Lu Zhang,[3] Clarence C. Lee,[4]

OPEN ACCESS Freely available online PLoS one

**Genetic Mapping and Exome Sequencing Identify Variants Associated with Five Novel Diseases**

RESEARCH ARTICLE

DIAGNOSTICS

**Rapid Whole-Genome Sequencing for Genetic Disease Diagnosis in Neonatal Intensive Care Units**

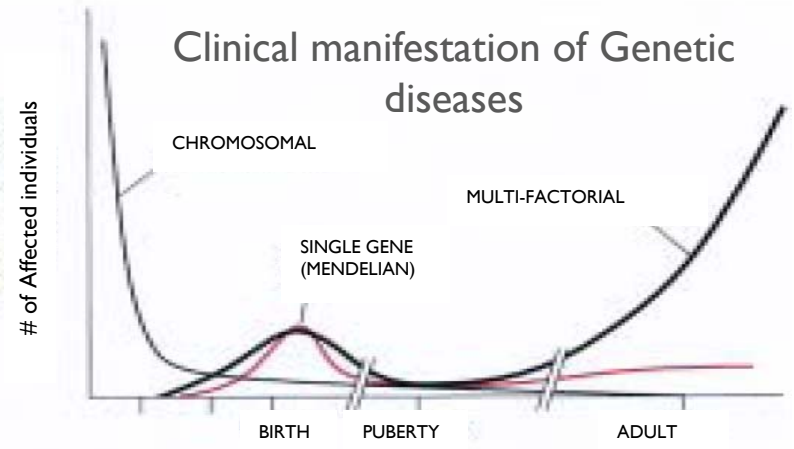
Carol Jean Saunders,[1,2,3,4,5]* Neil Andrew Miller,[1,2,4]* Sarah Elizabeth Soden,[1,2,4]* Darrell Lee Dinwiddie,[1,2,3,4,5]* Aaron Noll,[1] Noor Abu Alnadi,[4] Nevene Andraws,[3] Melanie LeAnn Patterson,[1,3] Lisa Ann Krivohlavek,[1,3] Joel Fellis,[6] Sean Humphray,[6] Peter Saffrey,[6] Zoya Kingsbury,[6] 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 Petrikin,[2,3] Kevin Peter Hall,[6] Stephen Francis Kingsmore[1,2,3,4,5†]

PARABASE GENOMICS

# Why Newborn Genomics?

- **Mendelian Diseases disproportionately affect Newborns**
  - ~3500 genetic diseases with molecular basis
  - >10% of NICU admissions are genetic
  - Current NBS tests limited to 29+ diseases
  - 2nd tier DNA testing to validate biochemical results

- **Advantage of NGS based DNA testing**
  - Find causal variants (rare/novel) in gene(s)
  - A 'universal' NGS approach avoids repeated, serial single gene testing
  - Current Sanger sequencing is expensive ($3-10K) and slow (3 months to 1 year)



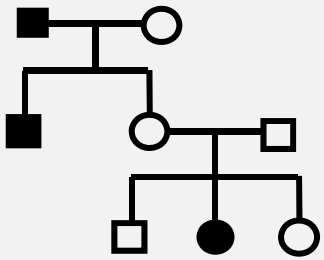

Gelehrter TD, Collins FS, Ginsburg D. Principles of Medical Genetics. 2nd ed. Baltimore, MD: Williams & Wilkins; 1998:1-42

NICU- Neonatal Intensive Care Unit
NBS-Newborn Screening
NGS-Next Generation Sequencing



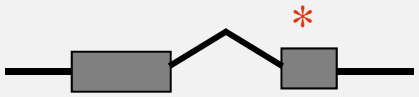
PARABASE GENOMICS

# Why Targeted (Exome) Sequencing for now?



Genomic DNA from Affected Individuals

NGS Sequencing




Causal Mutations in Exons/Target Regions

| Test Menu | Cost ($) | Throughput | Fold Efficiency |
|---|---|---|---|
| Whole Genome (Res.) | 7,666* | 1 | 1 |
| Exome (Res) | 1,200 | 15 | 95 |
| Neonate Panel (Clinical) | <1000 | 150 | >1140 |

- Majority of known disease-causing mutations in exons
- Exome = protein-encoding parts of genes
- Targeted NGS is Cost & Throughput Efficient

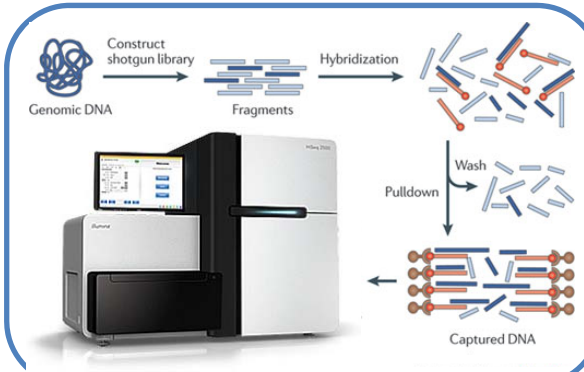
*Saunders et al., (2012) Rapid Whole Genome Sequencing for Genetic Disease Diagnosis in NICUs

PARABASE GENOMICS

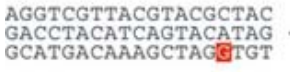
# Workflow for 2nd Tier Newborn Screening

**Sample Isolation** 2h

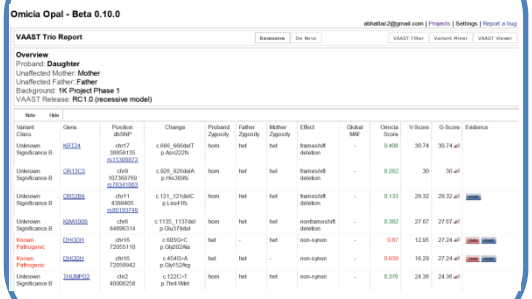
**DNA Capture & Sequencing** 92h




**Raw Data Management** 10h




**Analysis & Interpretation** 1h+




8 samples, 105 Hrs, <$10,000 = Real Neonatal Genomics!

PARABASE GENOMICS

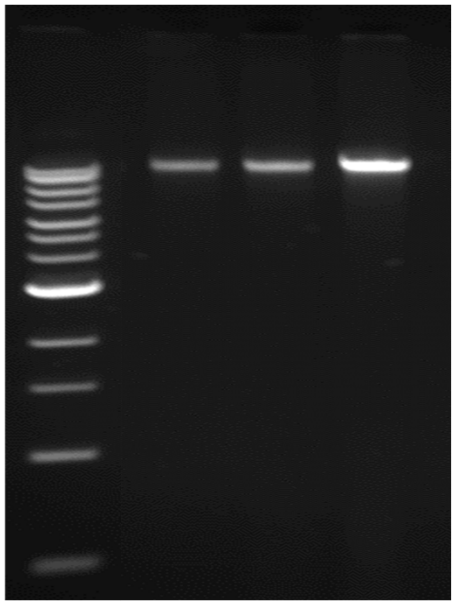
# Workflow for 2nd Tier Newborn Screening

**Sample Isolation**

2h

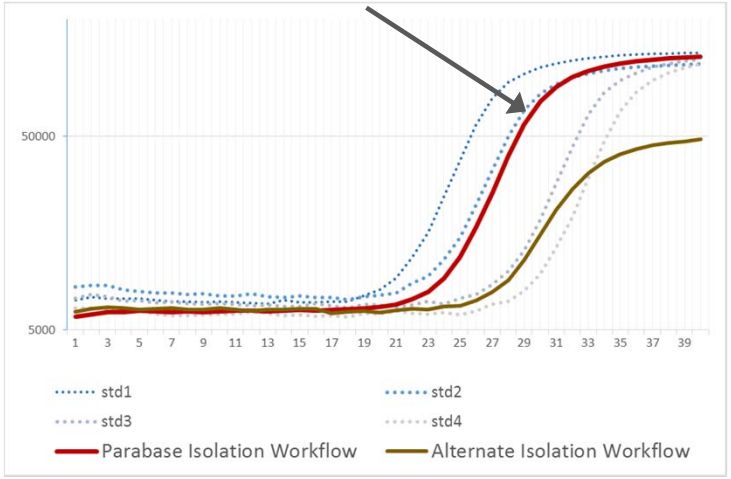
- High M.Wt. DNA
- PCR Amplifiable and NGS ready
- Processing Time ~ 2 hrs
- Amenable to automation

%DBS 25 50 100



Yield 400ng per Spot

More yield and DNA purity than alternate workflow




PARABASE GENOMICS

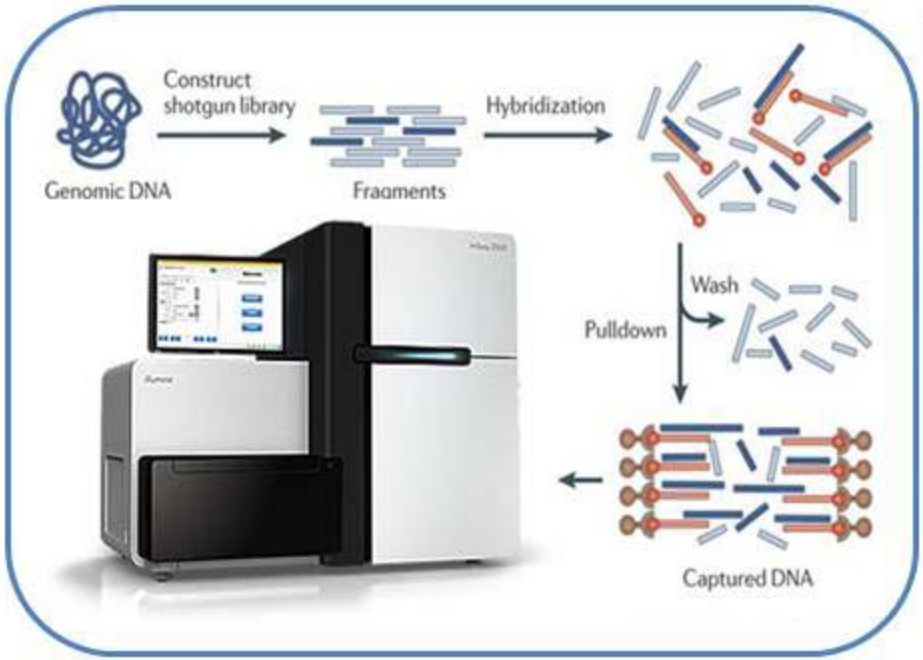
# Workflow for 2nd Tier Newborn Screening

**DNA Capture***
**& Sequencing****

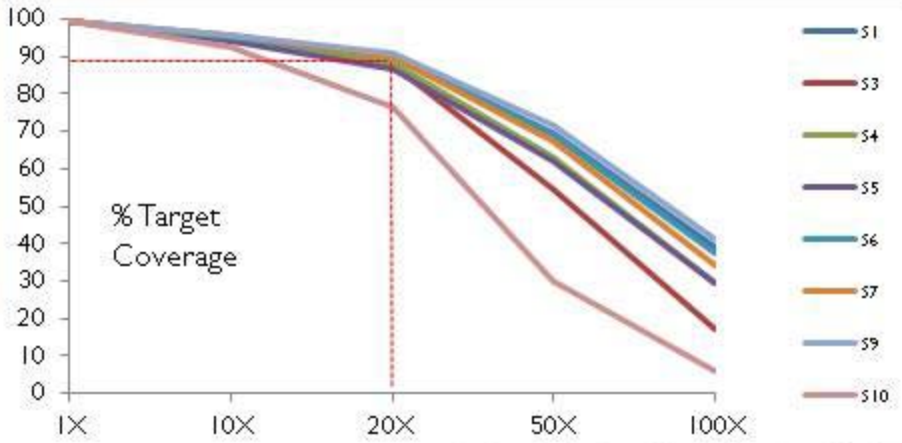
92h

** Room for improvement*
*** Process reduced from 19* days to 92 hrs*

- Library Prep ~ 5 hrs
- Capture ~ 66 hrs
- HiSeq 2500 PE 75 ~ 21 hrs
- Fast turnaround & High quality data




| | S1 | S3 | S4 | S5 | S6 | S7 | S9 | S10 |
|---|---|---|---|---|---|---|---|---|
| TOTAL READS (Mln) | 92 | 68 | 77 | 75 | 90 | 85 | 96 | 49 |
| MAPPED READS (Mln) | 89 | 58 | 72 | 72 | 86 | 79 | 91 | 37 |
| % READS MAPPED | 96 | 86 | 95 | 96 | 95 | 95 | 95 | 76 |
| READS ON TARGET (Mln) | 69 | 38 | 56 | 56 | 66 | 61 | 71 | 21 |
| % READS ON TARGET | 74 | 56 | 73 | 75 | 73 | 73 | 74 | 42 |



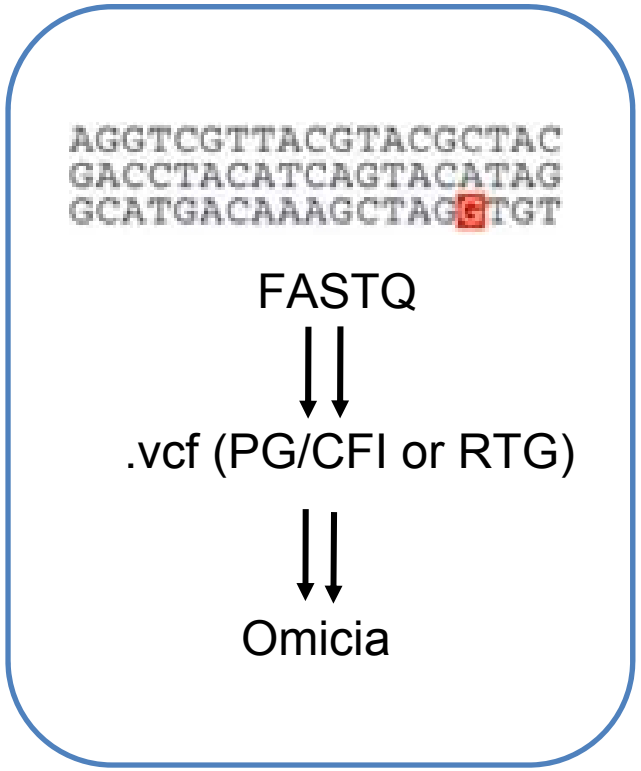

Data generated on Nimblegen Exome and HiSeq2500
Variant calling GATK2 (Broad Institute, MA) by CFI and Real Time Genomics
2 Known (MSUD and PA) & 8 blinded samples were provided by Clinic for Special Children, PA (S1-S10)
*Saunders et al., (2012) Rapid Whole Genome Sequencing for Genetic Disease Diagnosis in NICUs

PARABASE GENOMICS

# Workflow for 2nd Tier Newborn Screening

**Raw Data Management**
*(Mapping, alignment & variant calling)*

10h




- We have integrated and tested two fast data management and processing workflows
  - Clinical Future Inc. (CFI)-uses GATK2
  - Real Time Genomics (RTG)- proprietary
- All Variant (.vcf) files processed in ~10 hrs.
- High quality variant calls-PG/CFI pipeline

| Total Variants | 23Me | UCLA | PG/CFI |
|---|---|---|---|
| EdgeBio | 39116 | 25552 | 36458 |
| 23Me | | 30182 | 43417 |
| UCLA | | | 25378 |

| % Discordance | 23Me | UCLA | PG/CFI |
|---|---|---|---|
| EdgeBio | 0.36% | 0.39% | 0.27% |
| 23Me | | 0.31% | 0.35% |
| UCLA | | | 0.13% |



PARABASE GENOMICS

# Workflow for 2nd Tier Newborn Screening

**Analysis & Interpretation**
*(Annotation, interpretation & reporting)*

1h

- Tested two known cases: MSUD and PA
- Used CRADD for FP/FN testing
- Tested blinded samples, a set of 8 from 30

| NO | DISORDER | POPULATION | GENE | VARIANT |
|---|---|---|---|---|
| 1 | Phenylketonuria | Amish/Mennonite | PAH | 782 G>A |
| 2 | Maple Syrup Urine Disease | Mennonite | BCKDHA | 1312 T>A |
| 3 | Medium Chain Acyl-Co Dehydrogenase Deficiency | Mennonite | ACADM | 985 A>G |
| 4 | Glutaric Aciduria Type 1 | Amish | GCDH | 1262 C>T |
| 5 | Cystic Fibrosis | Amish/Mennonite | CFTR | 1522-24 del TTT |
| 6 | Severe Combined Immunodeficiency Disease | Amish | RAG1 | 2974 A>G |
| 7 | Severe Combined Immunodeficiency Disease | Mennonite | IL7R | 2 T>G |
| 8 | 21-Hydroxylase Deficiency | Amish | CYP21A2 | 518 T>A |
| 9 | 11-β-Hydroxylase Deficiency | Amish | CYP11B1 | 1343 G>A |
| 10 | 3-β-Hydroxy Steroid Dehydrogenase Deficiency | Amish | HSD3B2 | 35 G>A |
| 11 | Galactose Uridyl Transferase Deficiency | Amish | GALT | 563 A>G |
| 12 | Biotinidase Deficiency | Mennonite | BTD | 1459 T>C |
| 13 | Biotinidase Deficiency | Amish | BTD | 1368 A>C |
| 14 | Biotinidase Deficiency | Amish | BTD | 1330 G>C |
| 15 | Homocystinuria | Amish | MTHFR | 1129 C>T |
| 16 | Propionic Acidemia B | Amish/Mennonite | PCCB | 1606 A>G |
| 17 | Adenosine Deaminase Deficiency (SCID) | Amish | ADA | 646 G>A |
| 18 | Glutaric Aciduria Type 3 | Amish | C7orf10 | 895 C>T |
| 19 | 3-Methylcrotonylglycinuria 2β | Amish | MCCC2 | 295 G>C |
| 20 | 3-Methylcrotonylglycinuria 2β | Mennonite | MCCC2 | 518 ins T |
| 21 | 3-Methylcrotonylglycinuria 2β | Mennonite | MCCC2 | 687 A>C |
| 22 | Mevalonate Kinase Deficiency | Mennonite | MVK | 803 T>C |
| 23 | Mevalonate Kinase Deficiency | Mennonite | MVK | 1174 G>A |
| 24 | Galactose Uridyl Transferase Deficiency | Amish | GALT | 940 A>G |
| 25 | Phenylketonuria | Amish | PAH | 280-282 del ATC |
| 26 | Phenylketonuria | Mennonite | PAH | IVS 10-11 G>A |
| 27 | Phenylketonuria | Mennonite | PAH | IVS 12+1 G>A |
| 28 | Tyrosinemia Type 3 (Hawkinsinuria) | Mennonite | HPD | 85 G>A |
| 29 | Tyrosinemia Type 3 (Hawkinsinuria) | Mennonite | HPD | 479 A>G |
| 30 | Tyrosinemia Type 3 (Hawkinsinuria) | Mennonite | HPD | 1005 C>G |
| 31 | Methylmalonic/Homocystinuria, cblC Deficiency | Amish | MMACHC | 271 ins A |
| 32 | Medium Chain Acyl-CoA Dehydrogenase Deficiency | Mennonite | ACADM | IVS 4-30 A>G |

← Known

← Known

PARABASE GENOMICS

# Detecting Known Cases

## Identifying causal variants using population frequency and disease category as filters

| Type | Sample | Disease | Exome Variants | Protein Impact (PI) | PI+ Probably Damaging | PI+PD Hom.Read>4 MAF <5% | 126 Gene Panel Read >4 MAF<5% | 126 GP,Common Hom., Read >4 MAF<5% | Gene | Reads | Transcript Variant | Protein Variant | Status |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Amish* | whole blood | Propionic Acidemia | 71,261 | 10,127 | 1,013 | 11 | 19 | 2 | PCCB | 18 | c.1606A>G | p.Asn536Asp | Hom. |
| Amish* | DNA | Propionic Acidemia | 64,003 | 10,451 | 1,037 | 15 | 16 | 2 | PCCB | 5 | c.1606A>G | p.Asn536Asp | Hom. |
| Mennonite* | DNA | Maple Syrup Urine Disease | 68,703 | 10,217 | 1,031 | 19 | 19 | 3 | BCKDHA | 35 | c. 1312 T>A | p.Tyr438Asn | Hom. |
| Mennonite* | DNA | Mental Retardation NS | 69,946 | 10,329 | 1,086 | 21 | ND | ND | CRADD | 15 | c. 382G>C | p.Gly128Arg | Hom. |

## Validation of False Positive/ False Negatives by comparing different Methods

| CRADD Sample, Broad** | Exome Variants | Panel (*in silico*) |
|---|---|---|
| SYNONYMOUS | 11219 | 81 |
| MISSENSE | 10518 | 65 |
| NONSENSE | 91 | 0 |
| SMALL INDELS | 889 | 4 |
| INTRON, UTRs | 26083 | 232 |
| SPLICE SITE | 160 | 1 |


Data generated on Nimblegen Exome; variant calling GATK (Broad Institute, MA); Omicia (Emeryville, CA)

*Samples provided by Clinic for Special Children

** Puffenberger *et al.*, 2012. PLoS ONE 7(1): e28936. Agilent Exome using Broad Pipeline

PARABASE GENOMICS

# Detecting 8 blinded cases among 30

| NO | DISORDER | POPULATION | GENE | VARIANT | Sample |
|---|---|---|---|---|---|
| | | | | | |
| 1 | Phenylketonuria | Amish/Mennonite | PAH | 782 G>A | S10 |
| 2 | Maple Syrup Urine Disease | Mennonite | BCKDHA | 1312 T>A ← | Known |
| 3 | Medium Chain Acyl-Co Dehydrogenase Deficiency | Mennonite | ACADM | 985 A>G | S1 |
| 5 | Cystic Fibrosis | Amish/Mennonite | CFTR | 1522-24 del TTT | S3 |
| 7 | Severe Combined Immunodeficiency Disease | Mennonite | IL7R | 2 T>G | S6 |
| 9 | 11-β-Hydroxylase Deficiency | Amish | CYP11B1 | 1343 G>A | S9 |
| 11 | Galactose Uridyl Transferase Deficiency | Amish | GALT | 563 A>G | S5 |
| 12 | Biotinidase Deficiency | Mennonite | BTD | 1459 T>C | S7 |
| 15 | Homocystinuria | Amish | MTHFR | 1129 C>T | S4 |
| 16 | Propionic Acidemia B | Amish/Mennonite | PCCB | 1606 A>G ← | Known |

PARABASE GENOMICS

# All Samples 8/8 were correctly identified

| No. | Exome Variants | Protein Impact (PI) PG/CFI | PI RTG | %SNV >40 | Avg. Reads | PI+PD <5% AF | PI+PD <5% AF Hom | PI+PD >5 Read <5% AF Het. | 126 Neonatal Panel | 552 Hereditary panel | Gene | Reads | Transcript variant | Protein Variant | Zyg. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| S1 | 43373 | 13909 | 10468 | 95 | 102 | 10 | 2 | 10 | 1 | 4 | ACADM | 101 | c. 1085 A>G | p.Lys362Glu | Hom. |
| S3 | 43537 | 14142 | 13549 | 95 | 72 | 12 | 0 | 12 | 0(4) | 3(16) | CFTR | 43 | c.1521_1523del | p.Ser507del | Hom. |
| S4 | 43033 | 13968 | 10314 | 95 | 89 | 6 | 1 | 5 | 2 | 3 | MTHFR | 92 | c.1251C>T | p.Arg417Cys | Hom. |
| S5 | 42985 | 14123 | 10417 | 94 | 88 | 14 | 2 | 12 | 1 | 4 | GALT | 79 | c.564A>G | p.Gln188Arg | Hom. |
| S6 | 43759 | 14080 | 10585 | 95 | 99 | 10 | 0 | 10 | 0(3) | 0(24) | IL7R | 198 | c.3T>G | p.met1Arg | Hom. |
| S7 | 43706 | 14269 | 10749 | 95 | 97 | 14 | 0 | 14 | 0(4) | 0(14) | BTD | 74 | c.1460T>C | p.Trp487Arg | Hom. |
| S9 | 43244 | 14051 | 10580 | 96 | 108 | 7 | 3 | 4 | 2 | 4 | CYP11B1 | 57 | c.1343G>A | p.Arg448His | Hom. |
| | | | | | | | | | | | ADA | 66 | c.646G>A | p.Gly216Arg | Het. |
| | | | | | | | | | | | CDH23 | 64 | c.3309A>G | p.Asn1103Ser | Het. |
| S10 | 43356 | 14363 | 10266 | 92 | 48 | 16 | 0 | 16 | 4 | 15(61) | PAH | 33 | c.782G>A | p.Arg261Gln | Het. |
| | | | | | | | | | | | PAH | 35 | c.284_286del | p.(Ile95_Lys96del) | Het. |
| | | | | | | | | | | | MCC2 | 61 | c.295G>C | p.Glu99Gln | Het. |

Disease categories, allele frequency (AF) filters can quickly identify causal variants

Panels (*in silico*) based on neonatal diseases & symptoms can rapidly identify mutants

We were able to identify compound heterozygote (PAH) and carriers

Data was generated on variant calling of CFI and RTG and reviewed in Omicia Opal (Emeryville, CA)

- All Samples were provided by Clinic for Special Children
- Variant identification was faster on *in silico* neonatal panels than on entire Exome



PARABASE GENOMICS

# Summary

- Can be used as 2nd Tier Newborn testing including NICU
- Causal variants identified rapidly in targeted panels
- Targeted NGS panel can be used for screening monogenic disorders
- Targeted NGS is affordable unlike Whole Genome ($$$)
- Targeted NGS is efficient- Do >10 'samples to answer' in 100 Hrs.

For product, services and collaboration opportunities please contact us

www.parabasegenomics.com

info@parabasegenomics.com

PARABASE GENOMICS

# Acknowledgements


**Sample Isolation**

**Parabase Genomics**

Tanya Sokolsky, Ph.D.

Yael Katan-Khaykovich, Ph.D.

& colleagues

**DNA Sequencing**

**YCGA**

John Overton, Ph.D.

& colleagues

**Raw Data Management**

**Clinical Future Inc.**

Alexander Zaranek, Ph.D.

Adam Berrey

Jonathan Sheffi

**Real Time Genomics**

Francisco De-LaVega, D.Sc.

Brian Hilbush, Ph.D.

**Analysis & Interpretation**

**Omicia**

Martin Reese, Ph.D.

**NIH/NIDCD SBIR Phase I- 2013**




PARABASE GENOMICS

# Thank You

PARABASE GENOMICS