

Personalis, Inc.

CEO Overview

Board of Directors Meeting

May 8, 2012

CEO Overview

- Company retains focus on product for 1st revenue in 2012
 - Rich to present R&D planning & progress
- 4/30 ending cash above budget (Carol to review financials)
- Progress on move to new office (Carol to review – proposal to approve)
- Continued development of ethical infrastructure (Carol to review)
- Progress in recruiting (>60% have commercial experience)
- Publications & public presentations
- Four patent applications submitted based on earlier Stanford provisionals
- Market developments – Personalis perspective
- Accuracy development (Gold standard genome devt budget to approve)
- Personalis laboratory strategy (outsource & insource)

Publications & Upcoming Presentations

- Seminal paper by Mike Snyder & team published in Cell :
 - Ongoing clinical use of genome-scale sequencing
 - Not just a genome once in your life
- Multiple publications by Rong Chen
- JW invited to speak at CHI “The Clinical Genome” meeting
 - June 13, San Francisco
- JW invited to speak at Institute of Medicine meeting
 - Washington, DC, July

Market Developments

- Illumina / Roche
 - Financial pressure on Illumina may drive more aggressive approach to clinical potential
 - Illumina will be asked whether it is developing that opportunity as fast as Roche might
 - MiSeq moving to 510k; Cancer-gene pullout panel released
 - 2 x 250 reads now “summer” and only MiSeq
- Complete Genomics
 - Delivery times being quoted as 12-16 weeks
- Oxford Nanopore
 - No data seen yet, but tentative meeting in UK in July
- Life Ion Proton
 - Expect HapMap genome for analysis in the next month

Overview of Personalis Accuracy Program

Characterize errors

Analysis

Rates by variant type & Q-score
Underlying mechanisms
Medical interpretation impact
Regulatory impact

Methods

Comparison with literature
Studies on limited regions
Whole genome analysis:
 Single genome metrics
 Same person twice
 Family inheritance (MIE)
 Platform comparison
 Arrays
 ILMN / CGI / Ion T...
 Sanger capillary

Prioritize solutions

Bio-informatic approaches
Laboratory approaches
Content approaches

Informatic Solutions

Advances in 3rd party alg's

Aligner for > 200bp reads
New versions, current alg's
Other alg's (e.g. lobSTR)

Advances in our alg's:

SV's : Zygosity, accuracy
Combine SV's & SNP's
Broader use of Breakseq

New Personalis alg's

Alignment degeneracy
Genome-aware alignment
VNTR-specific

Advances in reference

Ethnic major-allele ref's
Iterative
Admixed
InDel's / SV's

Flag or Filter Out

For problems not yet solved

Lab, Content Solutions

Combine technologies

WGS + exome
Longer reads and/or inserts
Multi-vendor
Arrays (genotypes & SV's)
Electrophoretic sizing
Sanger validation

Content-based semi-custom

Array (genotypes & SV's)
Pullout

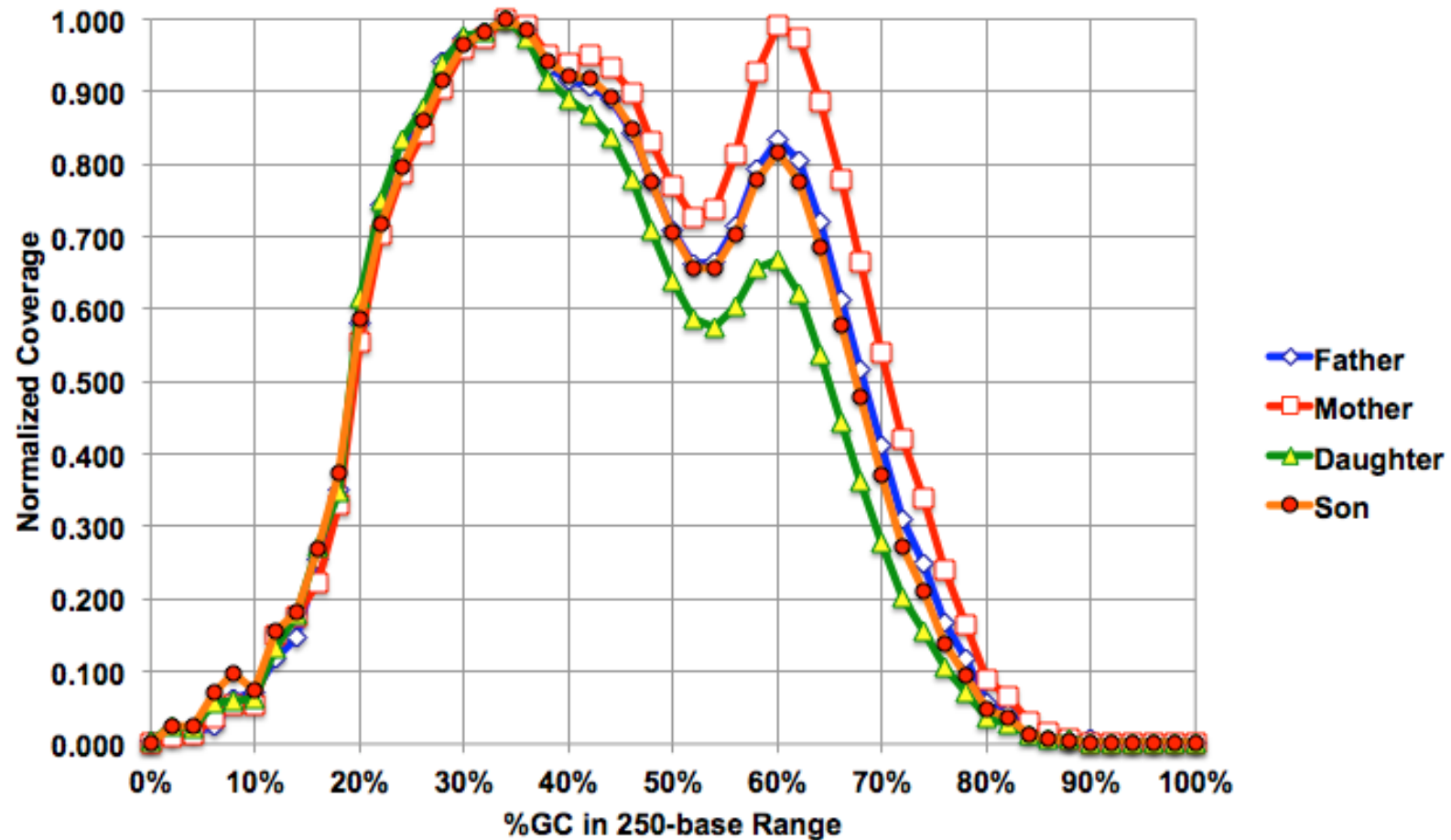
Custom sample preps

Problematic motifs
Degenerate regions
Phasing

Content

Gold standard genomes
Known healthy genomes

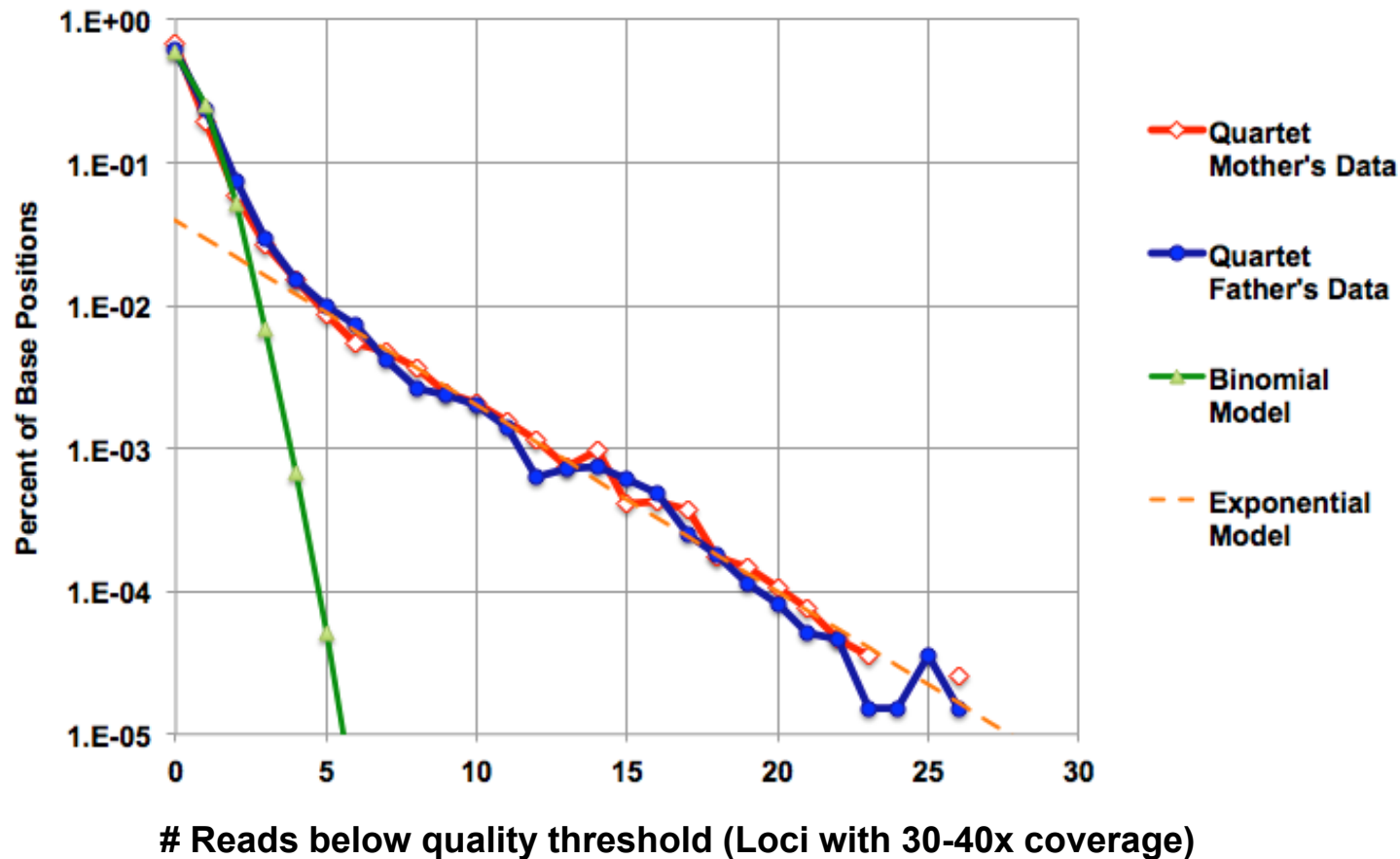
Above 50% GC, Quartet Genomes Have Substantially Different Coverage



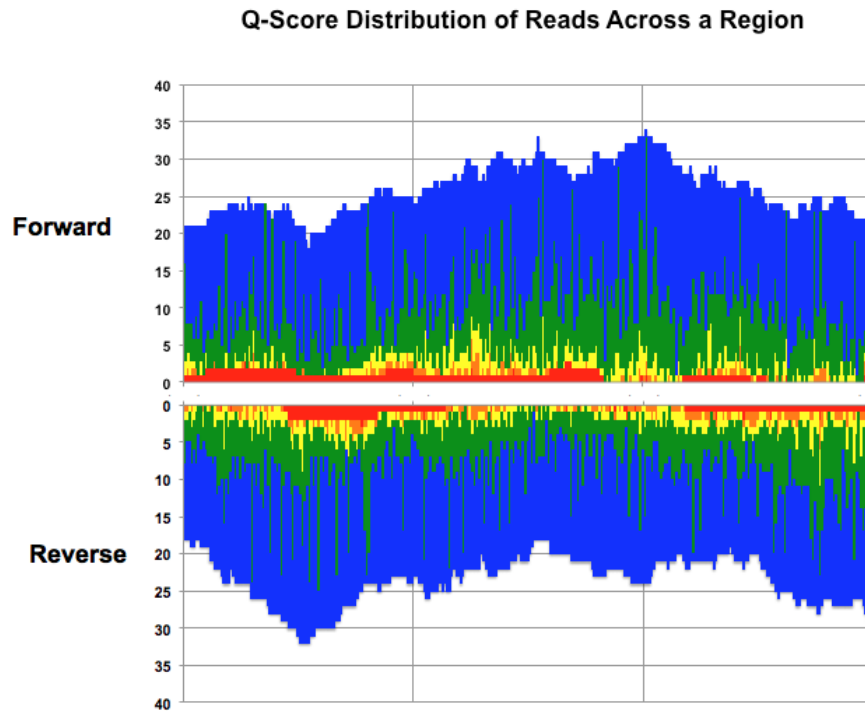
Example of process variation in genome sequencing

Four individuals had blood drawn together at the same hospital, sent to the same CLIA / CAP accredited lab, and sequenced on the same set of instruments.

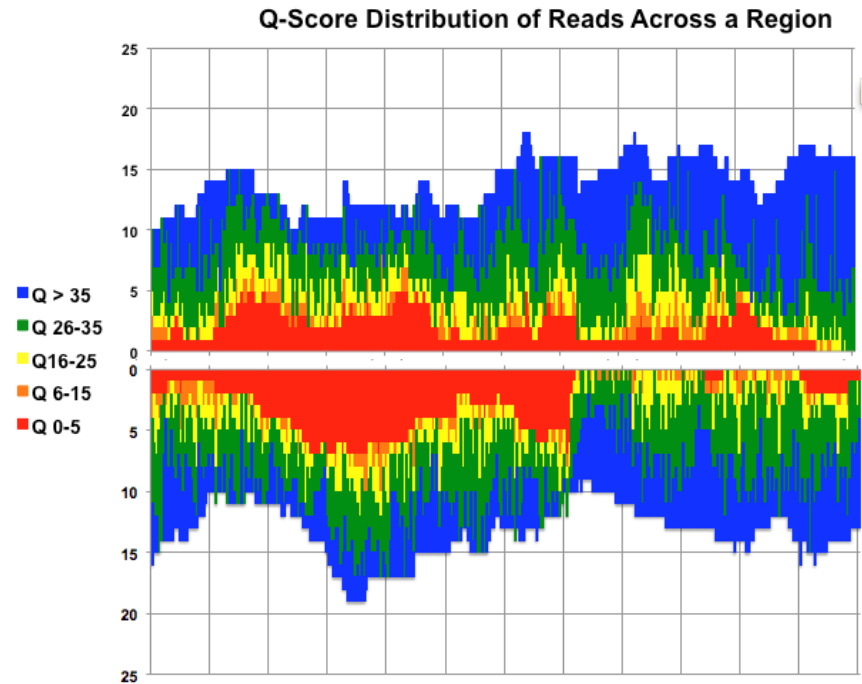
Occurrence of bases with very low raw quality scores deviates markedly from a random (binomial) model



Profile of systematic errors is remarkably similar between these two genomes



**Region with “normal”
sequence**



**Region with “problematic”
Sequence motif**

Tools to search for systematic errors across the genome, with their underlying mechanisms of action, will be followed by bio-informatic flags and (more importantly) bio-informatic and laboratory solutions.

Progression of Genome Data Sets for Accuracy Studies

Founder Genomes



Downloading



Custom Sequenced

West quartet (2009)

MIE analysis:

SNP (Dewey)

SV's (Manual, WG)

InDel (TBD)

Anecdotal error studies on
portions of the genome:

Phenomena

Error rates

Underlying mechanisms

GA IIx vs HiSeq (early 2011)

Snyder son/mother

Data behind publications:

Illumina vs CGI

Snyderome paper

Some inheritance validation

HapMap NA12878

Broad data set (late 2010):

Hi-Seq 2x100

10 PE insert lengths

Total coverage > 500

Same person "twice":

Same library

Different library

High coverage

SV validation (Snyder)

Fluor. Array data

HapMap genotypes

Some relatives' data

available

Korean PGP

40 HiSeq 2x100 genomes

CEPH1463

3 generations, 17 people

Mother is HapMap NA12878

To be "gold std"

All 17 sequenced by CGI

Test-bed for Personalis max
accuracy combination of
technologies

Use to assess other tech's:

Ion Proton

Oxford Nanopore

MiSeq 2x250

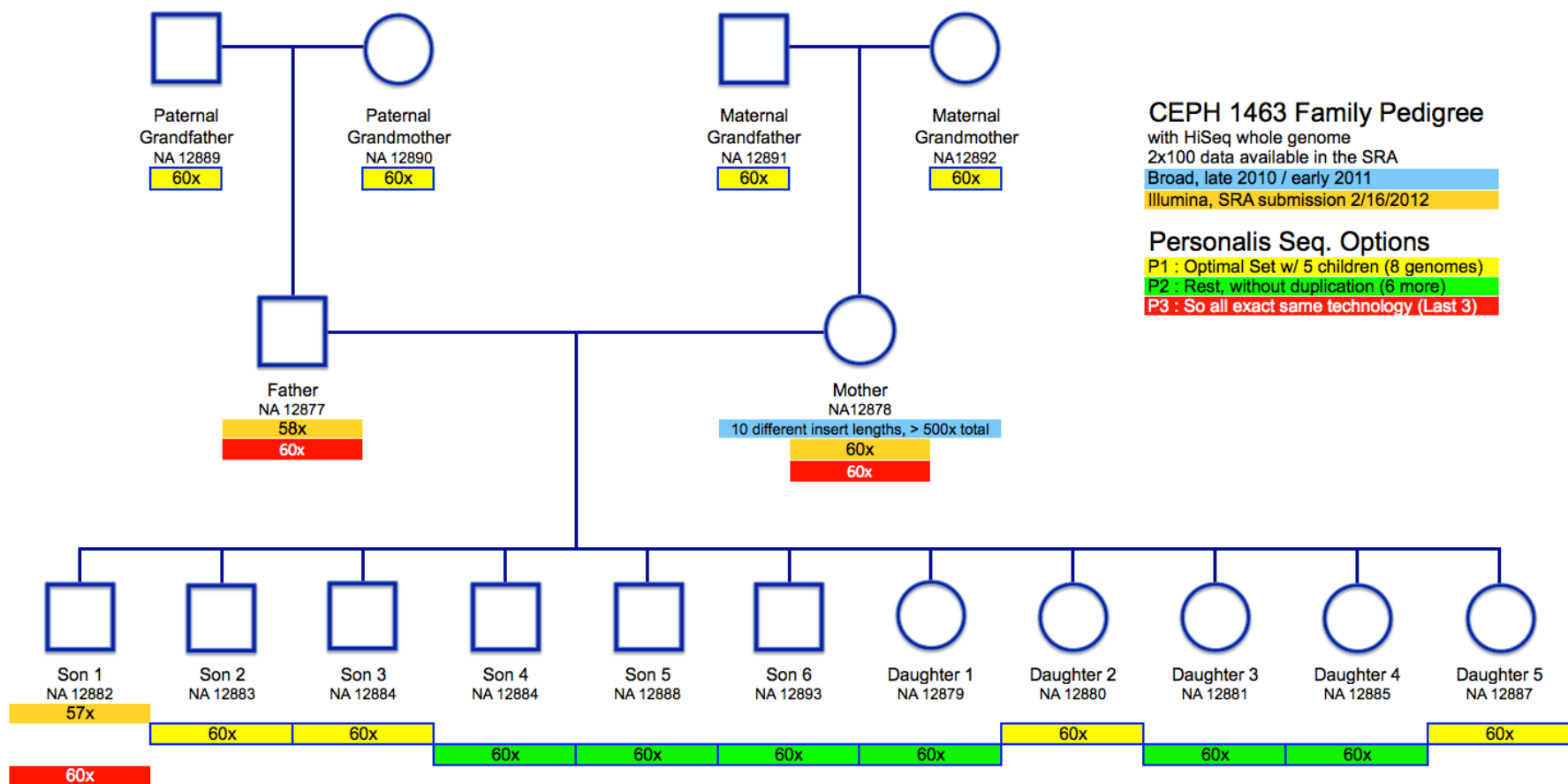
Anonymous samples

(ethics)

No Personalis relationship

Venter Genome

Only individual whole
genome on Sanger
technology



Segmentation of the 17-member pedigree by priority for sequencing

CEPH1463 Pedigree Analysis

CEPH1463 Pedigree Analysis		Percent of the Genome Over Which We Can See, Across the Data From All of the Children Included:						
		All Inheritance states		All Parental Alleles		All Genotypes possible from parents		
		Once	Twice	Once	Twice	Once	Twice	
All 11 Children		99.9%	99.9%	100.0%	96.8%	84.0%	32.0%	All 11 children Several optimal subsets with 5 of the 11 children
The combination of 5 children which maximizes % genome where all Inheritance States are represented at least once (Children # 1,3,7,8,11)		91.0%	36.8%	94.7%	35.1%	22.0%	0.0%	
As above, but at least twice (Children # 2,8,9,10,11)		87.4%	46.0%	89.1%	35.9%	20.6%	0.0%	
All I.S. at least once, but meeting the criteria from X chromosome analysis (Children # 2,3,5,6,11)		89.5%	36.8%	92.9%	38.6%	26.4%	0.0%	
As above, but using data from Son # 1 (recently sequenced by Illumina & submitted to the SRA) (Children 1,2,3,8,11)		88.3%	35.5%	92.6%	38.0%	24.9%	0.0%	

Mendelian Inheritance Error analysis can be confounded by genomic variation induced by the cell culture process. Being able to see each parental allele in two separate children should let us separate most of these errors from sequencing artifacts, letting us better identify them.

"GOLD STANDARD GENOME" PROJECT BUDGET

Amount is within budget but requires Board approval

Additional semi-custom work based on Personalis content may be proposed later in the year.

Samples de-identified before shipment to vendors

Whole genome sequencing

By Illumina Genome Network at Illumina	\$ 122,000
20 samples at 60x coverage	
17 CEPH1463 pedigree	
2 replicates	
Venter	
Complete Genomics Sequencing	\$ 10,000
2 Samples at 80x coverage	
Venter	
CEPH1463 Mother (replicate)	
Array analysis	\$ 5,000
Genotypes	
SV validation	
Karyotyping, exomes & other assays	\$ 20,000
TOTAL	\$ 157,000

Personalis Laboratory Strategy

- Laboratory goals:
 - Direct relationship with the customer
 - Quality control, sample to results (important for research and later clinical)
 - Position as “Value-Added Genome”, not “a bio-informatics company”
 - Offer a superior genome, by combining technologies from multiple companies
 - Develop superior sample preps, esp based on Personalis content
- Outsourcing:
 - Whole genome sequencing service pricing now < reagents to do it yourself
 - Avoid infrastructure costs (instruments, initial computation, staffing)
 - Turn-around time is a disadvantage (quotes at 12-16 weeks)
 - Other assays we can outsource:
 - Arrays (standard & semi-custom)
 - Karyotyping
 - Exomes, RNAseq
- Personalis lab
 - Receive, de-identify samples
 - QC, sample tracking
 - Ship aliquots to multiple vendors
 - Validations following primary analysis
 - Develop custom sample preps

Next : VP Finance report