From: Richard Chen
To: John West; Carol Tillis; Eliane Sousa
Subject: Final Board Slides
Date: Tuesday, May 8, 2012 10:17:29 AM
Attachments: board.05.08.12.v7.rtc.pptx

Here are my final board slides ready for print.

-Rich

R&D Update

Personalis, Inc.
Board Meeting 05.08.2012

Richard Chen, CSO

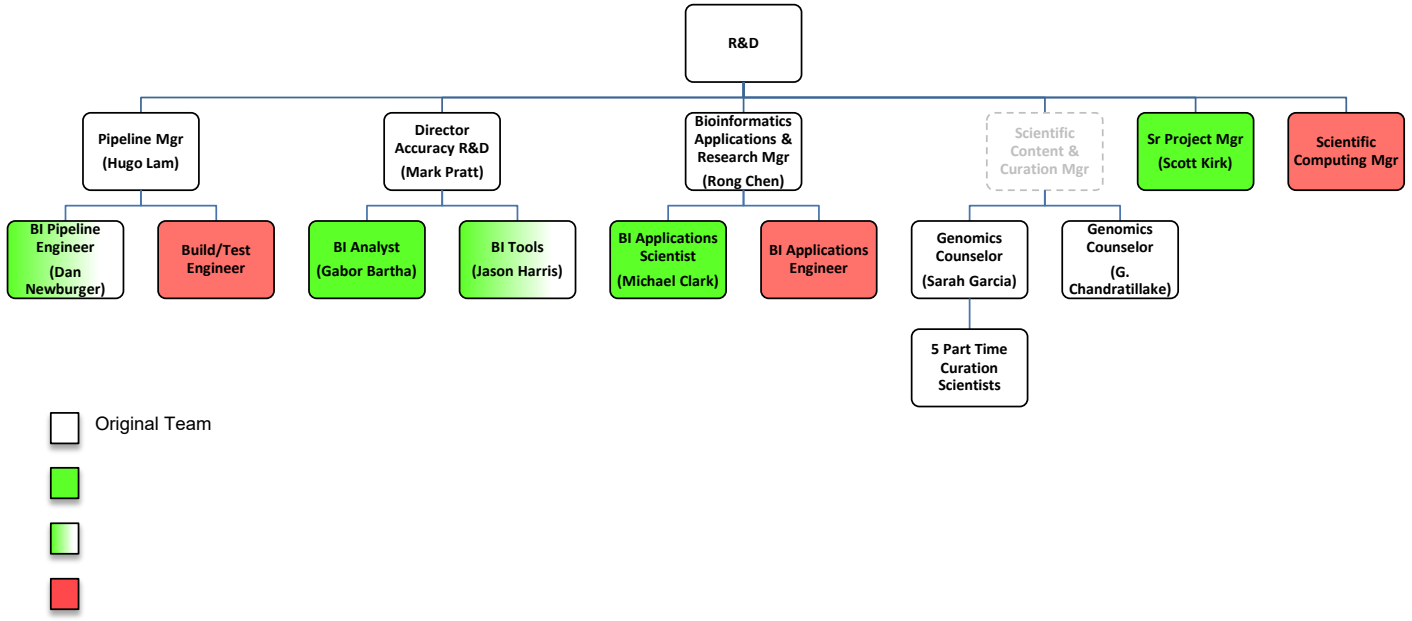
Overview

- Hiring
- Product Development
  - Genome Pipeline Service
  - Rare Variant Discovery Service
  - Genome Interpretation Service
- Content Development
- Research Collaborations
- Computing Infrastructure

R&D Hiring Update

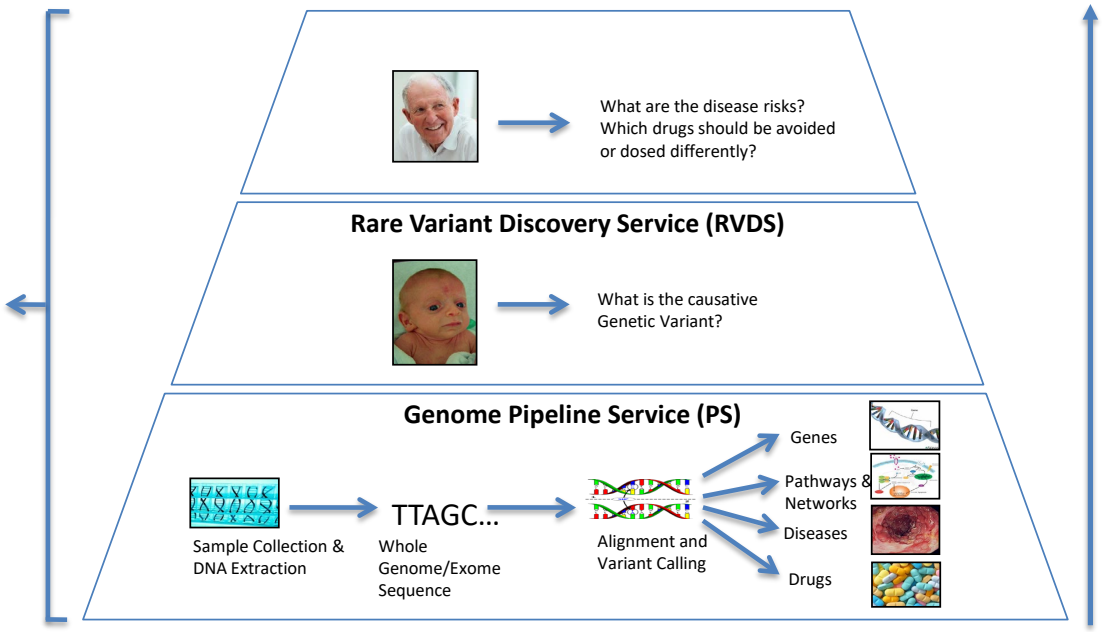
- New R&D Hires
  - Scott Kirk (Project Mgmt, started 4-1-12)
  - Michael Clark (Apps R&D, started 4-16-12)
  - Gabor Bartha (Accuracy R&D, started 5-1-12)
  - Jason Harris (Accuracy R&D, starting late May)
  - Daniel Newburger (Pipeline R&D, starting mid-June)

- Near Term Open Positions
  - Scientific Computing Mgr
  - Test/Build Engineer
  - Applications Engineer

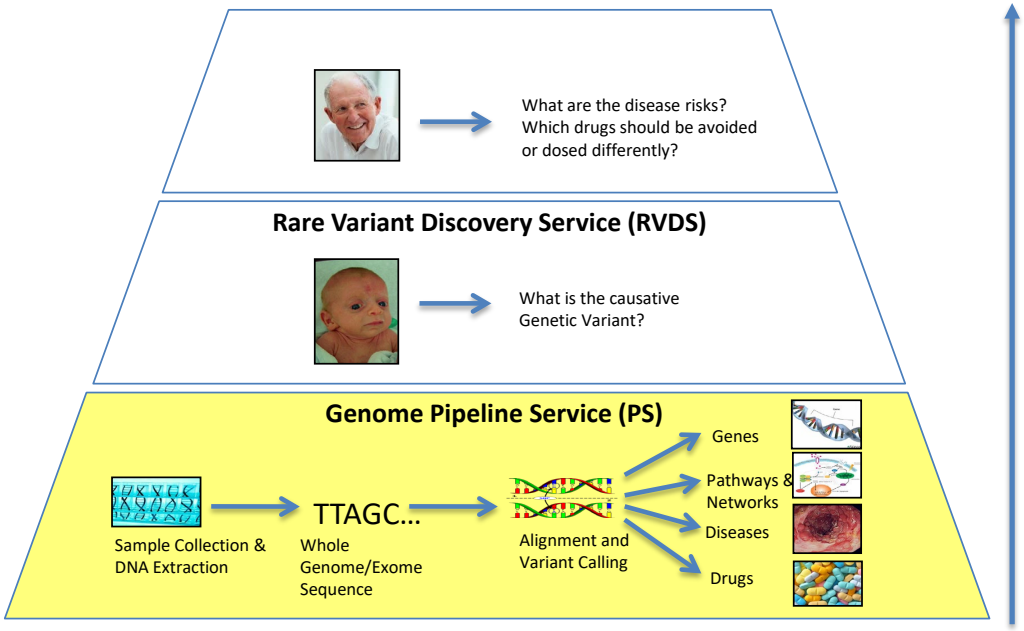


Product Roadmap

- Moving forward on all three services.
- Current dev focus is on Genome Pipeline Service.
- Spec'ing out and planning for RVDS and GIS.
- As resources allow, starting dev on long lead items for RVDS and GIS



Product Roadmap



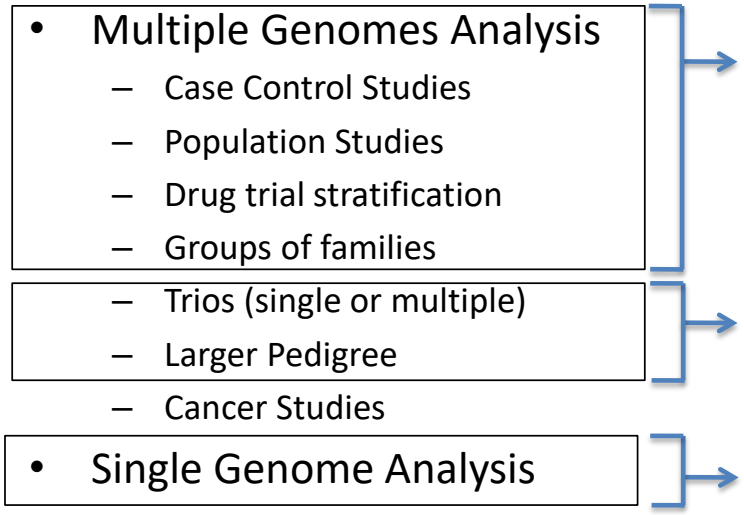
R&D Update: Genome Pipeline Service

- First draft of requirements specification completed but constantly refining
  - Developing Pipeline Service report to be compelling and differentiating
- Team currently focused on development of the prototype with the goal of testing early with collaborators
- Project plan completed with preliminary delivery date estimates, on track for:
  - Aug 2012: Prototype Product for Testing with Collaborators
  - Nov 2012: Estimated date for v1 Product for Beta Testing
- The accuracy effort is foundational for the Pipeline service (described by John earlier). Research risk may impact v1 schedule.
- Developing “upstream” ops capabilities to handle and process samples and different assays (Christian)
  - Need to match lab assay capabilities with downstream analytics

Pipeline Service: Primary Target Customer

Discovery Researcher Version: 0.1	Dr. Rick Discover
	"I'm very interested in applying these variants to my research." Background • Highly educated (PhD) or (MD/PhD) in a particular field and seeking to apply genomic information in their research. Looks to correlate genotypes with phenotypes to make new discoveries. • May have a great deal of software and computational skills, or only a little. Generally an expert at interpreting data sets. • Naturally questions results • May not have access to bioinformatics and computational resources to do large scale analysis, but wants flexible outputs that can be further analyzed in detail by themselves or someone on their team with the necessary skills. • Loves discovery and old-school straight black coffee.
Key Characteristics • Customer of Personalis service • Highly educated academic researcher • Ability to publish results is a key • If not computationally savvy has people who are. • Wants all of the data to interpret and further investigate	Goals • Want the whole data set to do further investigation on the results we provide. • Making new discoveries about variants, their associated phenotypes and publishing them. • Flexibility to ask a variety of questions of the data. • Needs confidence that results are correct, but comfortable making that determination from information presented to them. • Chasing false positives is costly and would like to avoid it. Usage patterns/Context: • Customer that orders whole genome or exome sequencing of a large number of samples using the Personalis service. • Want all of the data delivered to do further analysis

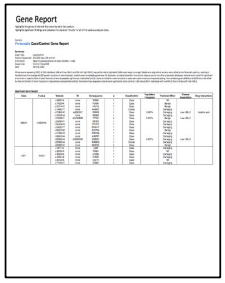
Pipeline Service: Target Use Cases

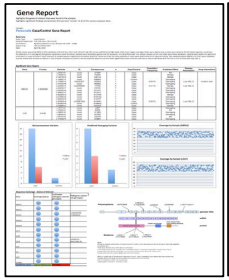


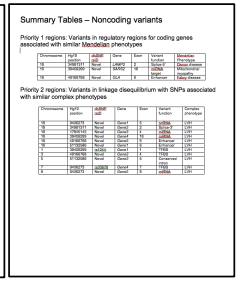
Pipeline Service

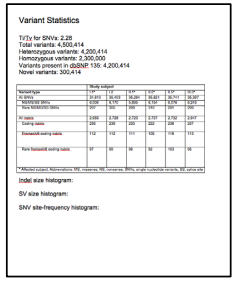
Pipeline Service: Report & Output

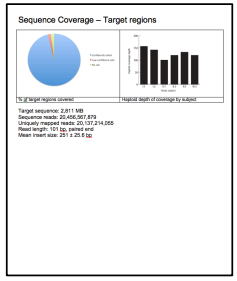
Detailed Reports





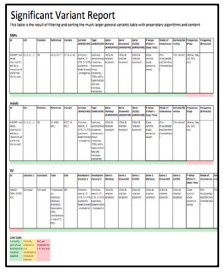


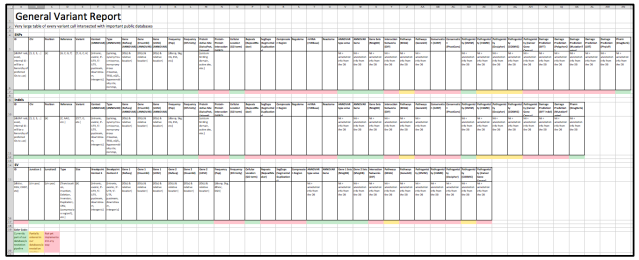


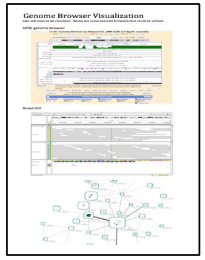


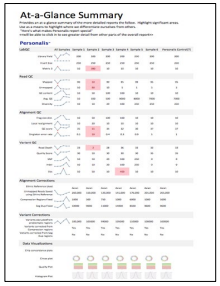
Biological Annotations

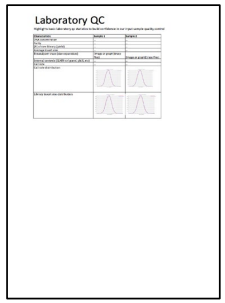
- Enable custom exploration and filtering of variants based on aggregated public DB annotations across all variants

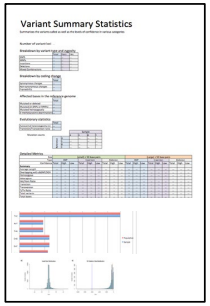


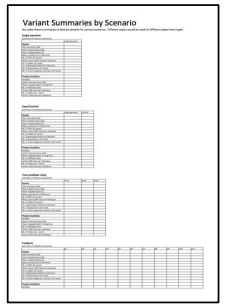












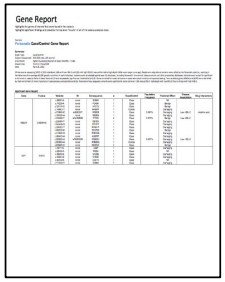
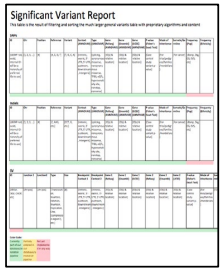
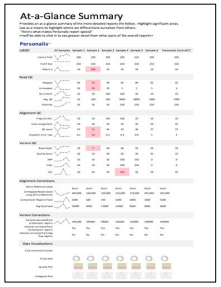
Personalis EX2042

Pipeline Service: Differentiation

		Commercial				Academic		Sequencing Product	
		Personalis	**Knome**	**DNANexus**	**RTG**	**Broad**	**BGI**	**Illumina**	**CG**
Platforms	**Illumina**	●	●	●	●	◒	●	●	○
	Complete Genomics	○	◒	◒	●	○	○	○	●
	Applied Biosystems	●	○	●	○	◒	●	○	○
Scalability	**Infrastructure**	Cluster	Cloud	Cloud	Cluster	Cluster (limited)	Cloud	Vendor	Vendor
	Data storage	○	○	●	○	○	○	○	○
	Integration	○	○	LIMS	○	○	○	○	○
Compatibility	**Input**	FASTQ/BAM	Raw	Raw	Raw	BAM	FASTQ	Raw	Raw
	Output	VCF/GFF	HGF	CSV	TSV	VCF	VCF	TSV	VAR
Variant detection	**Algorithms**	Multi/Open	Unknown	Unknown	Proprietary	Open	Open	Proprietary	Proprietary
	Mapping	●	●	●	●	◒	●	●	●
	SNP/Indel	●	●	●	●	●	●	●	●
	SV/CNV	●	○	○	◒	○	◒	○	●
	Annotation	●	◒	◒	○	○	◒	○	○
Availability	**Delivery**	End-user	Tools	Web	End-user	End-user	End-user	End-user	End-user
	Visualization	○	●	●	○	○	○	○	○
	Type	Service	Service	Service	Software	Software / Service	Software / Service	Service / Software	Service

➡ Our Key Differentiators will be Accuracy and Content

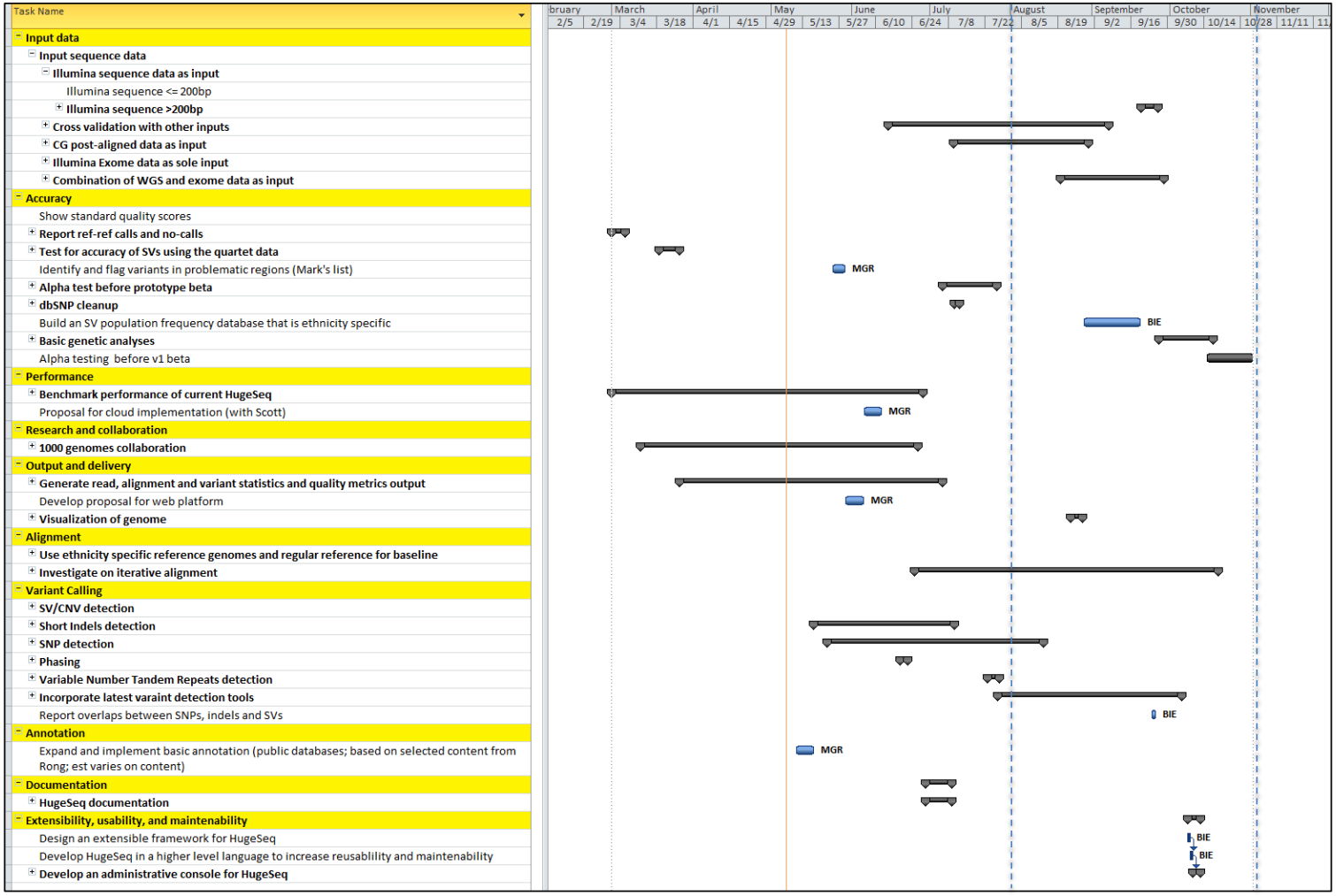
Pipeline Service: Differentiation

	Accuracy	Content	Other Features
	• Critical accuracy/quality statistics per variant and gene shown on report • Higher accuracy in pipeline enables FEWER candidate variants, and fewer false leads	• Use of proprietary content from Varimed, MendelDB, Regulome and PharmGKB to filter, analyze and prioritize most relevant variants in both coding and non-coding regions to the disease, pathway, drug or biological process of interest	• Ability to provide insights not just on SNPs and indels but also on SV's
	• Quality flags for problematic regions associated with each variant • Enable user determined quality thresholds for filtering rather than pre-set heuristics	• Annotation with content from a large of public databases (>15), more than available through other analysis services	• Integration with best of breed visualization tools such as UCSC, Broad IGV, Illumina Genome Studio, Spotfire, Cytoscape
	• QC metrics that highlight differences between standard pipeline vs Personalis pipeline results: • Rescue of variants using ethnicity specific references, better alignment methods, multiple assays • Flagging problematic variants (even those with high Q scores) and fixing them	• Use of proprietary databases of "medically important" variants to focus accuracy efforts on the most medically important regions	• Circos plot and other data visualizations for summarizing data that are "ready for publication"

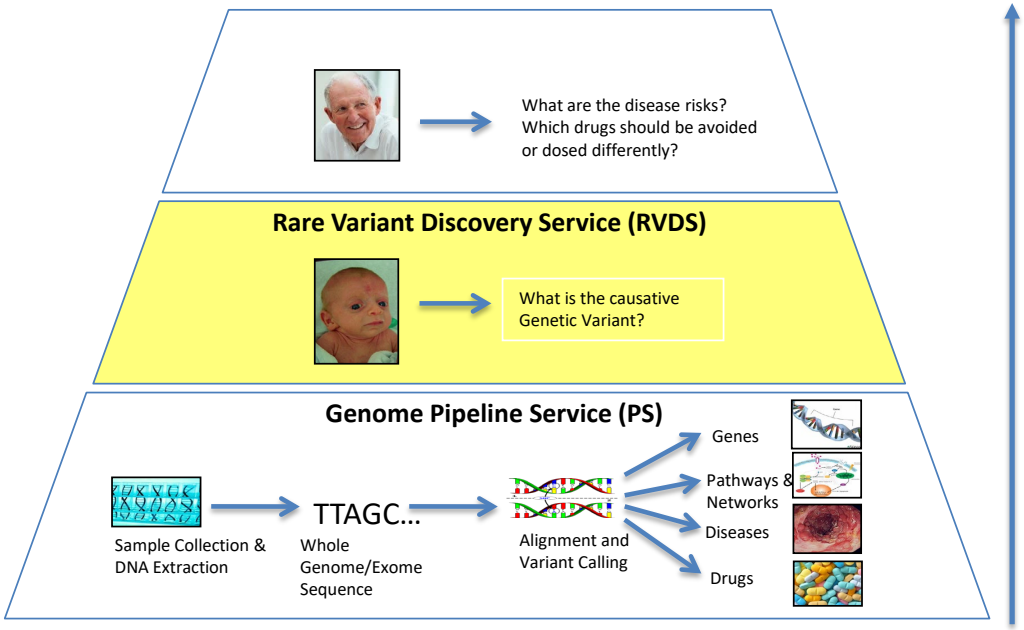
Pipeline Service: Prototype vs Version 1

	Prototype	Version 1
Goal	Test early with early friendly collaborators. Do enough product development to understand: •Overall Value proposition •Content/Functionality/Usability/Delivery •Rapid feedback for v1 product	Incorporate prototype feedback, generate revenue, scale •Revenue generation •"Whole product"
Assump	• Delivered as Service • Basic HugeSeq pipeline • Minimal new content/algorithms from where we are today • Basic quality controls, eliminate risk points from prototype	• More robust, scalable pipeline and service • New content and algorithms • Stricter quality controls for content and product • Hiring dependent
Inputs	• ILMN whole genome < 200bp	Multiple Assays including: •Combo of ILMN whole genome and exome •Exome as sole input •CG variants
Outputs	Basic report generation/data files as output including: •BAM files (alignment) and consensus sequence •Basic QC metrics •Variant calls with basic functional annotations •Ethnicity reference report with comparative analysis •Basic functional annotations •Biological annotations from public DBs	Expanded summary and quality reporting including •Expanded QC metrics •Differentiated visualization/graphs/reporting •Additional public DB annotation content
Accuracy Features	• 3 Ethnicity specific reference genomes • Flag variants in known problematic genome regions identified • Basic SNPs, indels, SV detection • Alleles reporting • Zygosity of largest indels	• Improved detection of ethnicity, handle admixture • Improved SV/CNV detection, zygosity detection • Overlaps between SNPs/Indels and SV's reported • Some of known problematic regions fixed

Pipeline Service: High Level Milestones



Product Roadmap



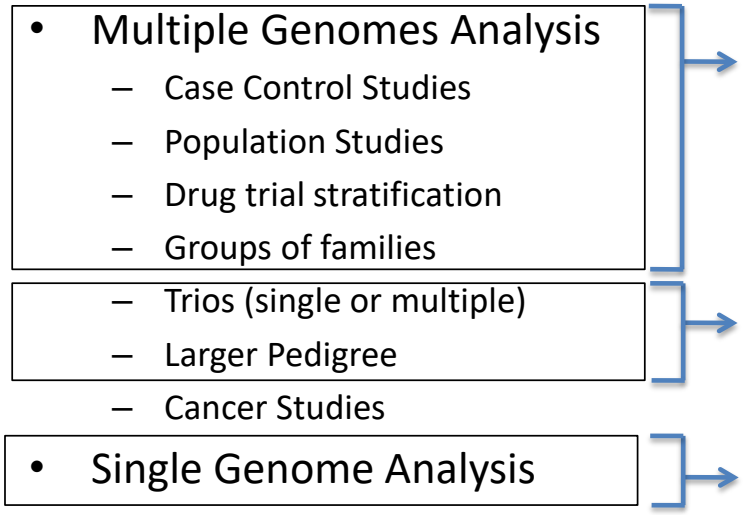
Rare Variant Discovery Service: Update

- Target customer is clinical researcher that has a family trio or large pedigree with a particular phenotype/disease of interest, under IRB
- Our service will take tissue samples or sequence as input and provide back a prioritized list of candidate variants that may be associated with the phenotype/disease of interest
- Completed first draft of functional spec and sample report, now in project planning and resourcing phase
  - Collaborated heavily with SAB (Dr. Rick Dewey and Dr. Euan Ashley) who are experts in this area to develop concept
- Overlapping resources with Pipeline Service development.
  - Early estimate for prototype is Dec 2012 timeframe
  - Already developing required content (MendelDB)
- Some reports/tools developed for Pipeline Service will be reusable

RVDS: Target Customer

Clinical Researcher Version: 0.1	Dr. Jane Hopkins
	"My goal is to use genetic information to uncover novel diseases, methods for diagnosing disease and treating disease" Background • Highly educated (MD/PhD) seeking to help specific patients that are being sequenced for diagnostic or therapeutic purposes. • Limited informatics experience • Generally an expert at interpreting data sets. • Typically does not have a team of bioinformaticians • Loves helping patients, identifying causes of disease Goals • Find novel diseases, methods for diagnosing disease, or treating disease using genetic data. • Using information to aid in making interpretations • Must have confidence that the results are correct • Publish Usage patterns/Context: • Customer that orders sequencing of individuals and small cohorts from Personalis. • Is delivered results from the service
Key Characteristics • Customer of Personalis service • Highly educated physician • Goal is helping patients • If not computationally savvy. • Wants confidence in the results being presented to them	

RVDS: Target Use Cases



RVDS Prototype Report

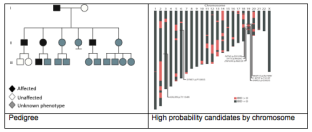
Personalis™

Disease Variant Discovery Service Results

Principal Investigator: Euan Ashley MRCP DPhil
Family ID: X00001

Summary

Disease: Hypertrophic Cardiomyopathy
Mode of inheritance: Autosomal dominant
Population prevalence: 1/500
Subjects sequenced: 6 (5 affected, 1 unaffected)
Sequencing platform: Illumina HiSeq 2000
Enrichment platform: none



Pedigree | High probability candidates by chromosome

Methods: Whole genome sequencing was performed in five affected individuals and one unaffected individual. Ninety-seven percent of bases in targeted regions were confidently genotyped, and mean haploid coverage depth was 46x. Variants were filtered according to co-segregation with disease and prioritized by variant type, frequency in ethnically concordant population data, and functional annotation.

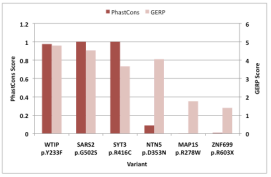
Consanguinity: none reported and no abnormal autozygosity identified.

Interpretation: No variants previously implicated in Hypertrophic Cardiomyopathy were identified. Four missense SNVs were identified in six tier 1 regions (exons) that cosegregated with disease, all in heterozygous state. These six variants have not been previously seen in exome sequencing of 5,400 subjects unselected with regard to Mendelian phenotypes, 1,094 subjects undergoing whole genome sequencing (1000 genomes phase 1 project), and are not catalogued in dbSNP. One variant, *LAMP2* p.Y233F, disrupts a highly conserved nucleotide in a gene that is highly expressed in myocardial tissue; *LAMP2* has been implicated in the closely related disorder, Danon disease. The other three candidate variants occur in genes (*SARS2, GLA*) with as-yet undetermined roles in hypertrophic cardiomyopathy, though they have been implicated in similar Mendelian disorders. These results are provided for research purposes only and are not intended for clinical use.

Frederick Dewey, MD — Sarah Garcia, PhD, MS

High-probability candidates

Chromosome	Hg19 position	dbSNP rsID	Gene	Exon	Variant function
19	34981311	Novel	*LAMP2*	2	Missense
19	39408299	Novel	*SARS2*	16	Missense
19	49166768	Novel	*GLA*	6	Missense
19	51132586	Novel	*GLA*	6	Missense



Variant Statistics

Ti/Tv for SNVs: 2.28
Total variants: 4,500,414
Heterozygous variants: 4,200,414
Homozygous variants: 2,300,000
Variants present in dbSNP 135: 4,200,414
Novel variants: 300,414

	Study subject					
Variant type	1*	2	3.1*	3.2*	3.3*	3.4*
All SNVs	34,810	35,403	35,264	35,821	35,741	36,397
NS/MS/SS SNVs	6,006	6,170	5,995	6,154	6,078	6,243
Rare NS/MS/SS SNVs	297	305	289	310	281	299
All indels	2,699	2,728	2,723	2,737	2,732	2,847
Coding indels	235	239	233	222	238	237
Frameshift coding indels	112	112	111	105	118	113
Rare frameshift coding indels	97	99	96	92	103	96

* Affected subject. Abbreviations: MS, missense; NS, nonsense, SNVs, single nucleotide variants, SS, splice site

Indel size histogram:

SV size histogram:

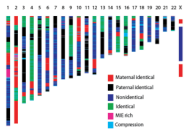
SNV site-frequency histogram:

Sequence Coverage – VIP genes

Genes previously implicated in the disease	Coverage – exons	% of pathogenic variants confidently genotyped	Pathogenic variants not confidently genotyped
ACTC			
CAV3			THR64SER [dbSNP:rs1168407?9]
GLA			
LAMP2			
MYBPC3			
MYH7			ARG249GLN [dbSNP:rs?2187133]
MYL2			
MYL3			MET149VAL [dbSNP:rs1048937348]
PRKAG2			
TNNC1			
TNNI3			
TNNT2			
TPM1			

Confidently called | Low confidence call | Not called

Genotype Quality Summary



Variants in compression regions: 1.6%
Variants in Mendelian inheritance abnormality-rich regions: 1.5%
Estimated per base genotype error rate*: 8.5 x 10^{-6}

*Estimated by Mendelian inheritance error and state consistency errors per confident genotype

Summary Tables – Coding variants

Priority 1 regions: Variants in coding genes associated with similar Mendelian phenotypes

Chromosome	Hg19 position	dbSNP rsID	Gene	Exon	Variant function	Amino acid change	Variant priority	Mendelian Phenotype
19	34981311	Novel	*LAMP2*	2	Missense	Y233F	2	Danon disease
19	39408299	Novel	*SARS2*	16	Missense	G502S	2	Mitochondrial myopathy
19	49166768	Novel	*GLA*	6	Missense	D353N	2	Fabry disease
19	51132586	Novel	*GLA*	6	Missense	R416C	2	Fabry disease
1	39408299	rs1564…	*Gene1*	1	Synonymous	--	3	AMPK cardiomyopathy
3	49166768	Novel	*Gene2*	4	Synonymous	--	3	LVNC
5	51132586	Novel	*Gene3*	5	Synonymous	--	3	LVNC
7	9408273	rs236078	*Gene4*	7	Synonymous	--	3	LVNC
8	9408273	Novel	*Gene5*	8	Synonymous	--	3	LVNC

Priority 2 regions: Variants in linkage disequilibrium with coding SNPs associated with similar complex phenotypes

Chromosome	Hg19 position	dbSNP rsID	Gene	Exon	Variant function	Amino acid change	Variant priority	Complex phenotype
19	9408273	Novel	*Gene1*	5	Nonsense	R223X	1	LVH
19	34981311	Novel	*Gene2*	2	Missense	F568DF	2	LVH
19	17845145	Novel	*Gene3*	4	Missense	R278W	2	LVH
19	39408299	Novel	*Gene4*	16	Missense	G502S	2	LVH
19	49166768	Novel	*Gene5*	6	Missense	D353N	2	LVH
19	51132586	Novel	*Gene1*	6	Missense	R416C	2	LVH
1	39408299	rs1564…	*Gene1*	1	Synonymous	--	3	LVH
3	49166768	Novel	*Gene2*	4	Synonymous	--	3	LVH
5	51132586	Novel	*Gene3*	5	Synonymous	--	3	LVH
7	9408273	rs23678	*Gene4*	7	Synonymous	--	3	LVH
8	9408273	Novel	*Gene5*	8	Synonymous	--	3	LVH

Summary Tables – Noncoding variants

Priority 1 regions: Variants in regulatory regions for coding genes associated with similar Mendelian phenotypes

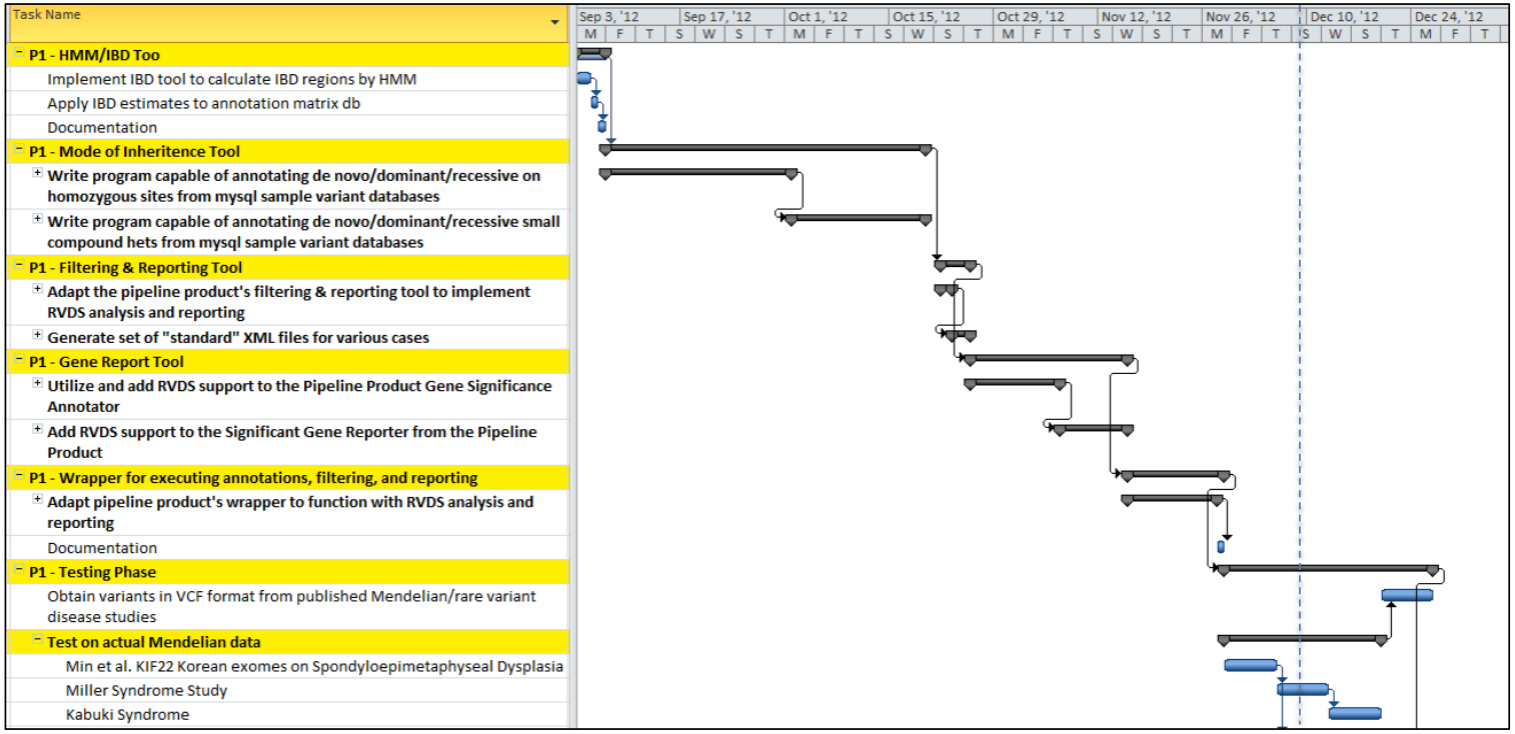
Chromosome	Hg19 position	dbSNP rsID	Gene	Exon	Variant function	Mendelian Phenotype
19	34981311	Novel	*LAMP2*	2	Splice-5'	Danon disease
19	39408299	Novel	*SARS2*	16	mRNA target	Mitochondrial myopathy
19	49166768	Novel	*GLA*	6	Enhancer	Fabry disease

Priority 2 regions: Variants in linkage disequilibrium with SNPs associated with similar complex phenotypes

Chromosome	Hg19 position	dbSNP rsID	Gene	Exon	Variant function	Complex phenotype
19	9408273	Novel	*Gene1*	5	ncRNA	LVH
19	34981311	Novel	*Gene2*	2	Splice-3'	LVH
19	17845145	Novel	*Gene3*	4	mRNA	LVH
19	39408299	Novel	*Gene4*	16	mRNA	LVH
19	49166768	Novel	*Gene5*	6	Enhancer	LVH
19	51132586	Novel	*Gene1*	6	Enhancer	LVH
1	39408299	rs1564…	*Gene1*	1	TFBS	LVH
3	49166768	Novel	*Gene2*	4	TFBS	LVH
5	51132586	Novel	*Gene3*	5	Conserved intron	LVH
7	9408273	rs23678	*Gene4*	7	TFBS	LVH
8	9408273	Novel	*Gene5*	8	mRNA	LVH

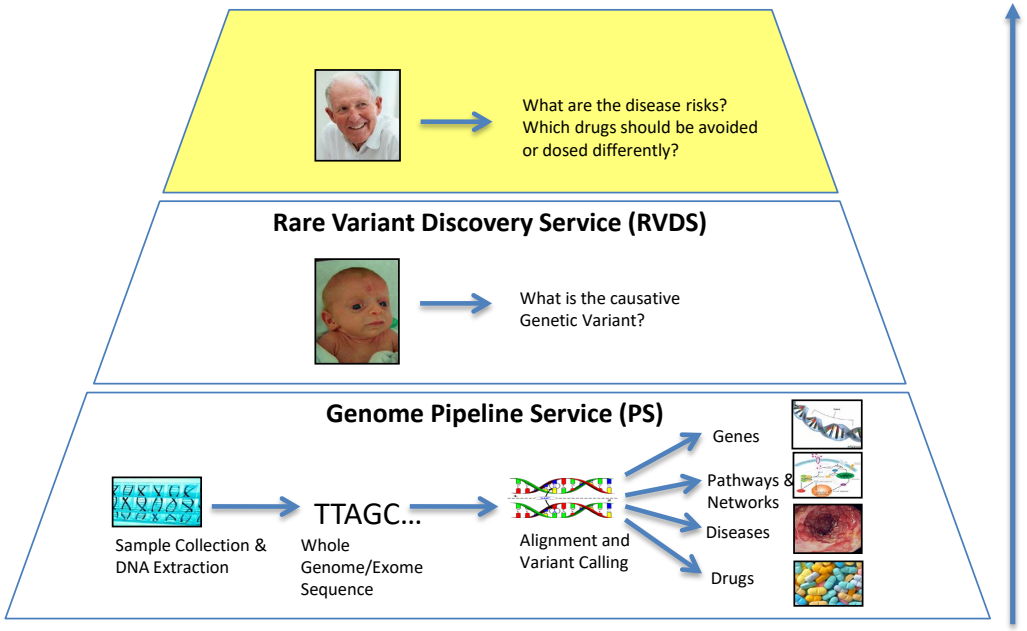
Personalis EX2042

RVDS: High Level Milestones



Personalis EX2042

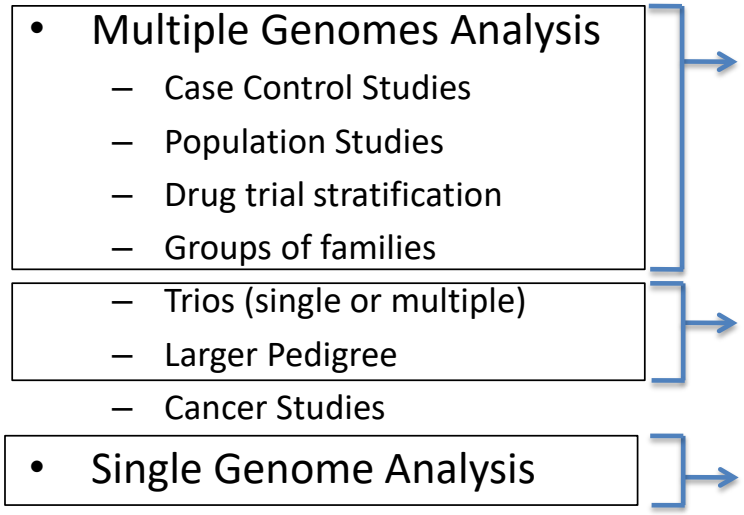
Product Roadmap



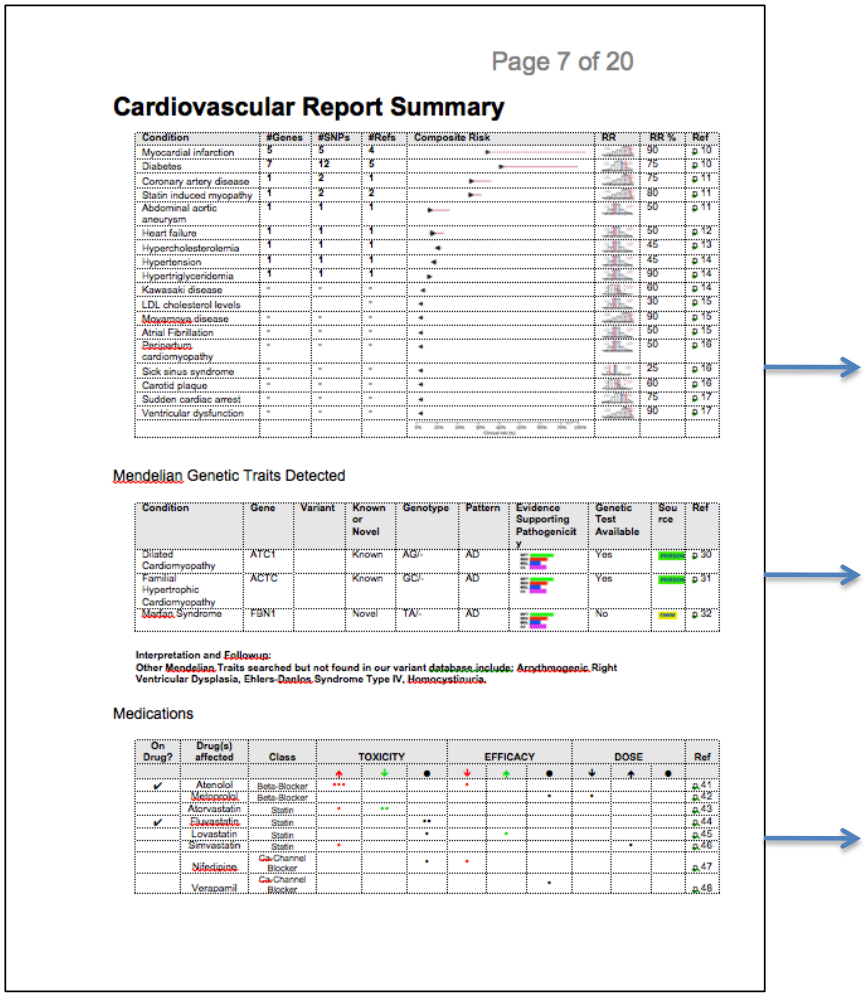
GIS: Target Customer

Clinical Researcher Version: 0.1	Dr. Jane Hopkins
	"My goal is to use genetic information to uncover novel diseases, methods for diagnosing disease and treating disease" Background • Highly educated (MD/PhD) seeking to help specific patients that are being sequenced for diagnostic or therapeutic purposes. • Limited informatics experience • Generally an expert at interpreting data sets. • Typically does not have a team of bioinformaticians • Loves helping patients, identifying causes of disease Goals • Find novel diseases, methods for diagnosing disease, or treating disease using genetic data. • Using information to aid in making interpretations • Must have confidence that the results are correct • Publish Usage patterns/Context: • Customer that orders sequencing of individuals and small cohorts from Personalis. • Is delivered results from the service
Key Characteristics • Customer of Personalis service • Highly educated physician • Goal is helping patients • If not computationally savvy. • Wants confidence in the results being presented to them	

GIS: Target Use Cases



Genome Interpretation Service



Page 7 of 20

Cardiovascular Report Summary

Condition	#Genes	#SNPs	#Refs	Composite Risk	RR	RR %	Ref
Myocardial infarction	5	5	4			90	p 10
Diabetes	7	12	9			75	p 10
Coronary artery disease	1	2	1			75	p 11
Statin induced myopathy	1	2	2			80	p 11
Abdominal aortic aneurysm	1	1	1			50	p 11
Heart failure	1	1	1			90	p 12
Hypercholesterolemia	1	1	1			45	p 13
Hypertension	1	1	1			45	p 14
Hypertriglyceridemia	1	1	1			90	p 14
Kawasaki disease	"	"	"			60	p 14
LDL cholesterol levels	"	"	"			30	p 15
Moyamoya disease	"	"	"			90	p 15
Atrial Fibrillation	"	"	"			90	p 15
Peripartum cardiomyopathy	"	"	"			90	p 16
Sick sinus syndrome	"	"	"			25	p 16
Carotid plaque	"	"	"			60	p 16
Sudden cardiac arrest	"	"	"			75	p 17
Ventricular dysfunction	"	"	"			90	p 17

Mendelian Genetic Traits Detected

Condition	Gene	Variant	Known or Novel	Genotype	Pattern	Evidence Supporting Pathogenicity	Genetic Test Available	Source	Ref
Dilated Cardiomyopathy	ATC1		Known	AG/-	AD		Yes		p 30
Familial Hypertrophic Cardiomyopathy	ACTC		Known	GC/-	AD		Yes		p 31
Marfan Syndrome	FBN1		Novel	TA/-	AD		No		p 32

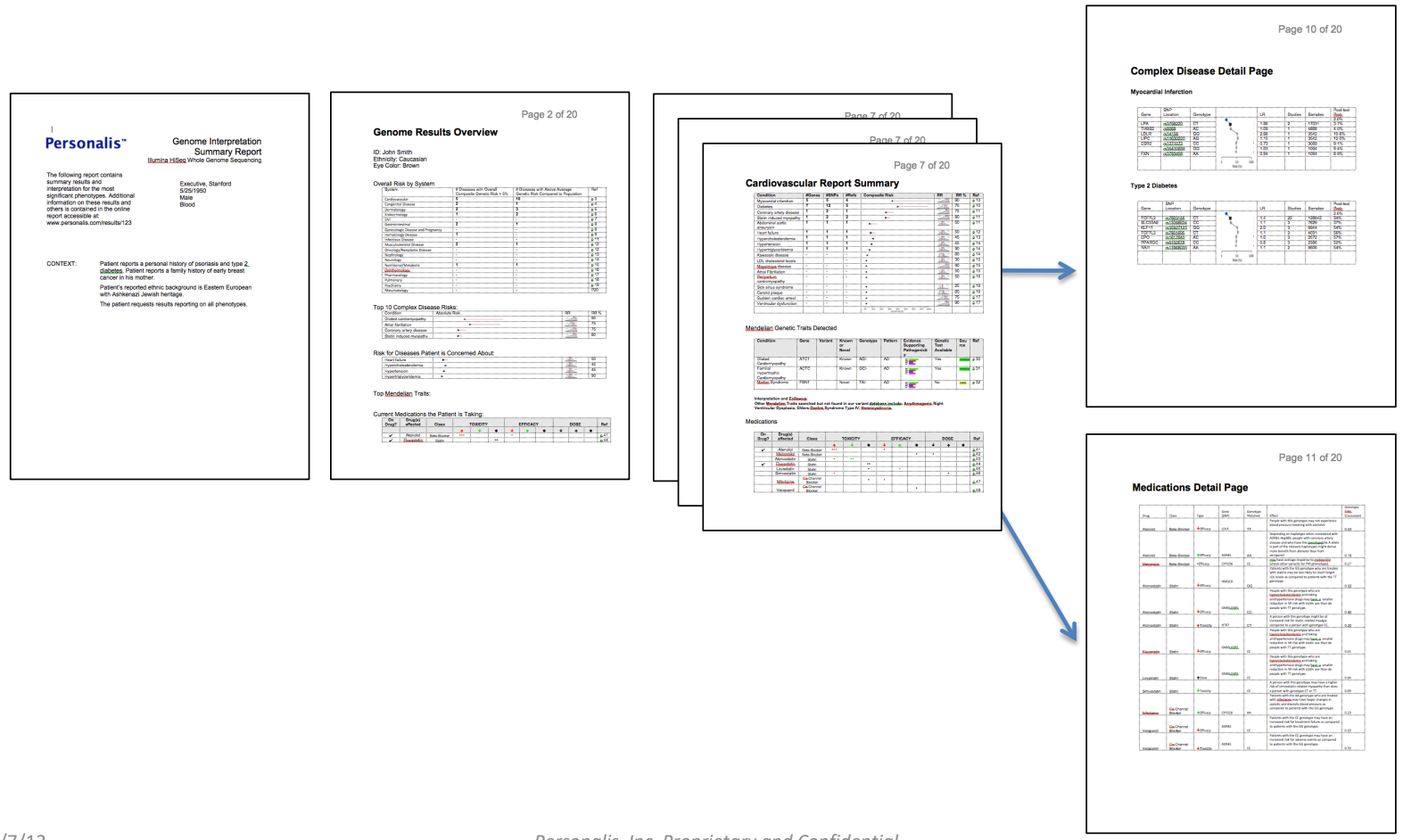
Interpretation and Followup:
Other Mendelian Traits searched but not found in our variant database include: Arrythmogenic Right Ventricular Dysplasia, Ehlers-Danlos Syndrome Type IV, Homocystinuria.

Medications

On Drug?	Drug(s) affected	Class	TOXICITY			EFFICACY			DOSE			Ref
✔	Atenolol	Beta-Blocker	***			*						p 41
	Metoprolol	Beta-Blocker						*	*			p 42
	Atorvastatin	Statin	*		**							p 43
✔	Pravastatin	Statin			**							p 44
	Lovastatin	Statin			*	*						p 45
	Simvastatin	Statin	*							*		p 46
	Nifedipine	Ca-Channel Blocker			*	*						p 47
	Verapamil	Ca-Channel Blocker						*				p 48

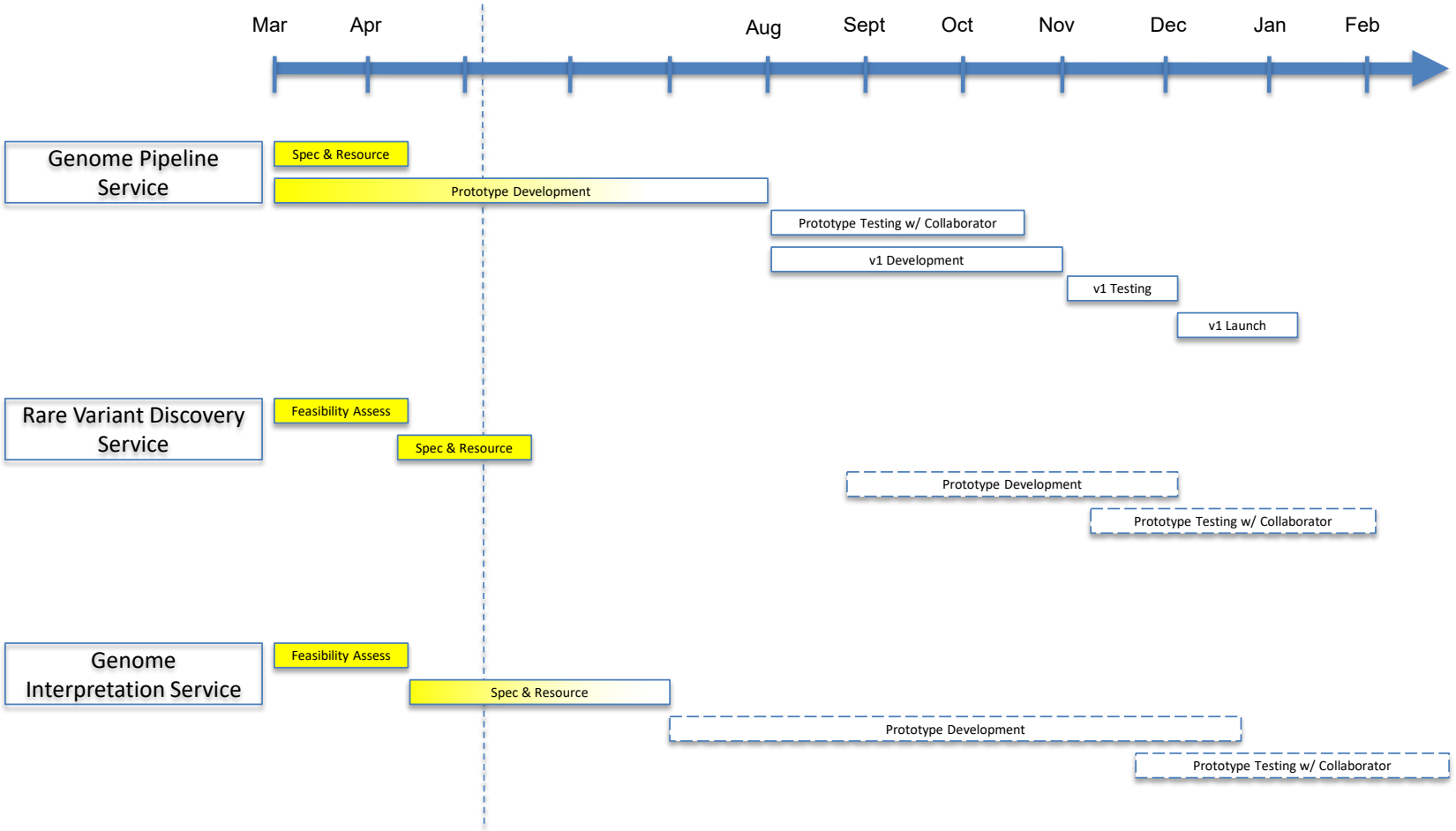
Genome Interpretation Service

- Now creating functional specification and project plan
  - Very Preliminary estimates for prototype: Dec 2012-Mar 2013



Personalis EX2042

Product Dev: Summary of Estimated Timelines



Personalis EX2042

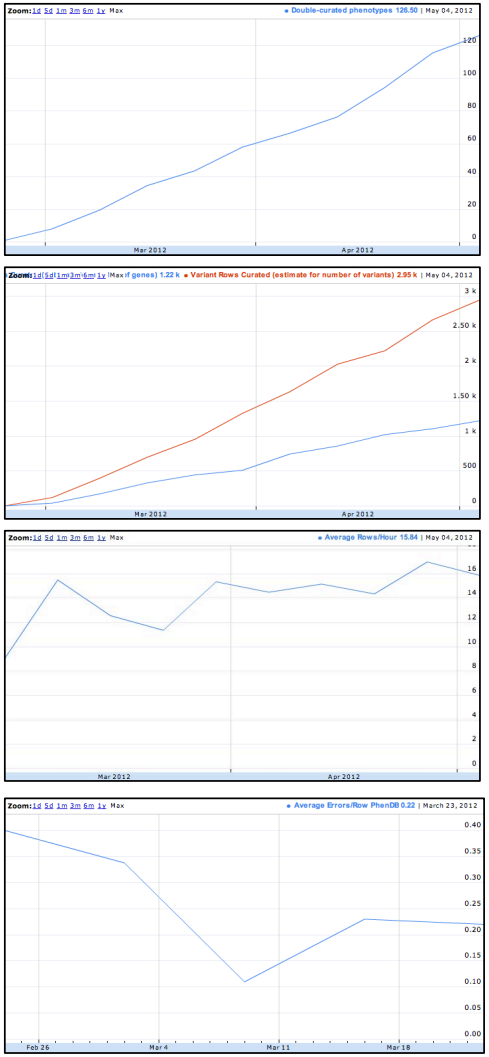
Product Dev: Summary of Resource Allocation

	Accuracy R&D **Mark Pratt** **Gabor Barth** Jason Harris (Start 5/27)	**Pipeline R&D** **Hugo Lam** Daniel Newburger (Start 7/15)	**Apps R&D** **Rong Chen** **Michael Clark**	**Content R&D** **Sarah Garcia** **Gemma Chandratillake** **PT Curators**
Pipeline Service				
Rare Variant Discovery Service				
Genome Interpretation Service				
Research Collaborations				
Ethics Board				

Content Development

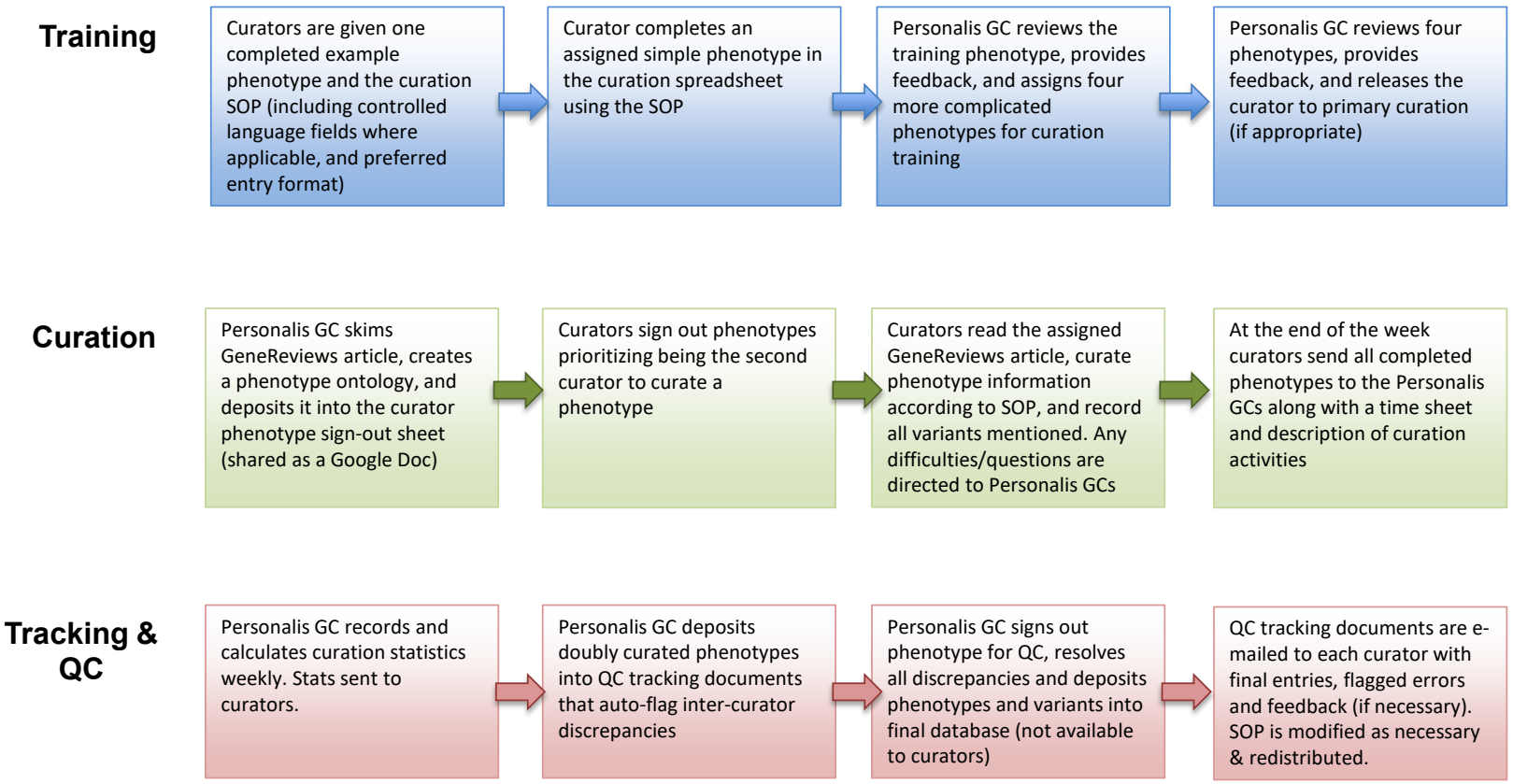
MendelDB Dashboard

- High quality curated DB of Mendelian phenotypes and their variants
  - Managed by Sarah Garcia and Gemma Chandratillake, both Genetics Ph.D. & Genetic Counselors
- Critical content to support Rare Variant Discovery Service and Genome Interpretation Service
- Ontological structure
- PT curation scientist workforce
  - 5 hired, top 20% of applicants tested
  - Majority work 8-15 hrs/week
  - Trained on protocol and real data for 2 weeks
  - Have been curating GeneReviews for 2 months
- PT curators receive weekly feedback on errors made, productivity
  - Also get feedback on average for the entire group
  - Rapid improvement
- Stringent QC process involving double curation and final QC check
- Expect to finish all (~560) GeneReviews by Aug 2012, corresponding to >200 common Mendelian disease phenotypes
- Will then continue phenotype curation and variant curation from other sources
- Flexible workforce for other curation tasks in the future



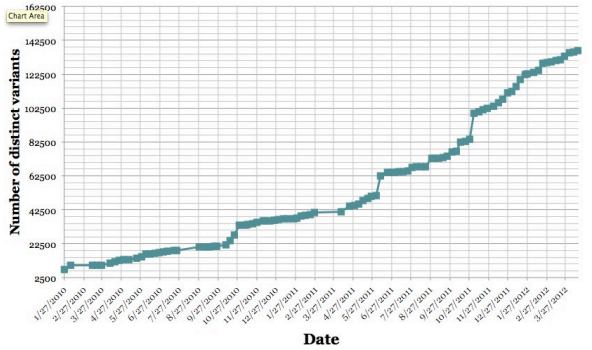
Raw error rate (# errors/row) (prior to QC)

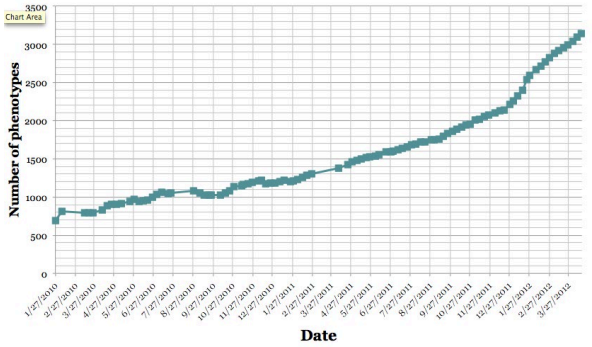
Mendel DB Curation Process

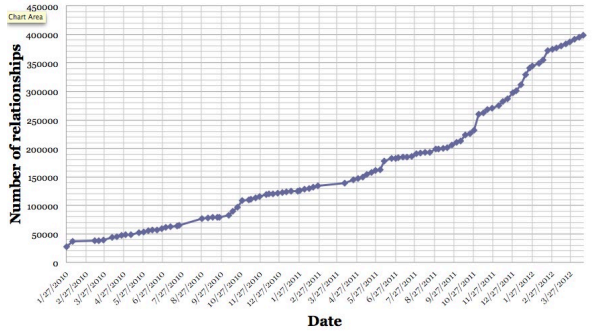


VarimedDB Dashboard

- Varimed is a DB of variant to disease relationships
- Curated by Optra workforce in India
- Future Issues
  - Continued process/quality review
  - Assess key “missing content” needs
  - Assess cost/benefit of incremental content







Personalis EX2042

RegulomeDB

- Licensed from Stanford
- A database of known and predicted regulatory elements
  - Manually curated regions that have been experimentally characterized in regulation
  - Transcriptional factor binding sites from CHIP-seq across a diverse set of cell types
  - DNA hypersensitivity site from Dnase-seq across over 100 cell types

Score	Supporting data	# SNPs
1a	eQTL + TF binding + matched TF motif + matched DNase Footprint + DNase peak	352
1b	eQTL + TF binding + any motif + DNase Footprint + DNase peak	2,566
1c	eQTL + TF binding + matched TF motif + DNase peak	85
1d	eQTL + TF binding + any motif + DNase peak	85
1e	eQTL + TF binding + matched TF motif	54
1f	eQTL + TF binding / DNase peak	34,694
2a	TF binding + matched TF motif + matched DNase Footprint + DNase peak	36,841
2b	TF binding + any motif + DNase Footprint + DNase peak	353,726
2c	TF binding + matched TF motif + DNase peak	16,640
3a	TF binding + any motif + DNase peak	302,597
3b	TF binding + matched TF motif	14,949

Regulome takes a list of variants and assign a score to each variant.
1a-1f: highly likely to be involved in regulation, 39,433 SNPs
2a-2c: likely to be involved in regulation, 407,797 SNPs

Supercentenarian Collaboration

- Research Collaboration with
  - Dr. Stuart Kim, Stanford Prof of Genetics
  - Also involved are Lee Hood from ISB and L. Stephen from Gerontology Research Group
- Goal: To uncover genetic factors that permit extreme longevity
  - Do supercentenarians have fewer overall risk alleles for common diseases that help them avoid disease?
  - Do supercentenarians have protective alleles that allow them to avoid aging/disease?
  - Do supercentenarians have CNVs that provide protection against or promote aging/disease?
- Details:
  - >=110 yrs old, 71 such people in the world, most female
  - ~17 supercentenarians (>=110 years old) have been sequenced (CG)
  - Analyzing variants using our content/algorithms
  - Completed IRB approval for Personalis involvement in project
  - Finalizing collaboration agreement with Stanford
- Benefits:
  - Potentially high profile, impact publication
  - Will showcase the power of our content from Varimed
  - Research experience with genomes from extremely healthy, long lived people

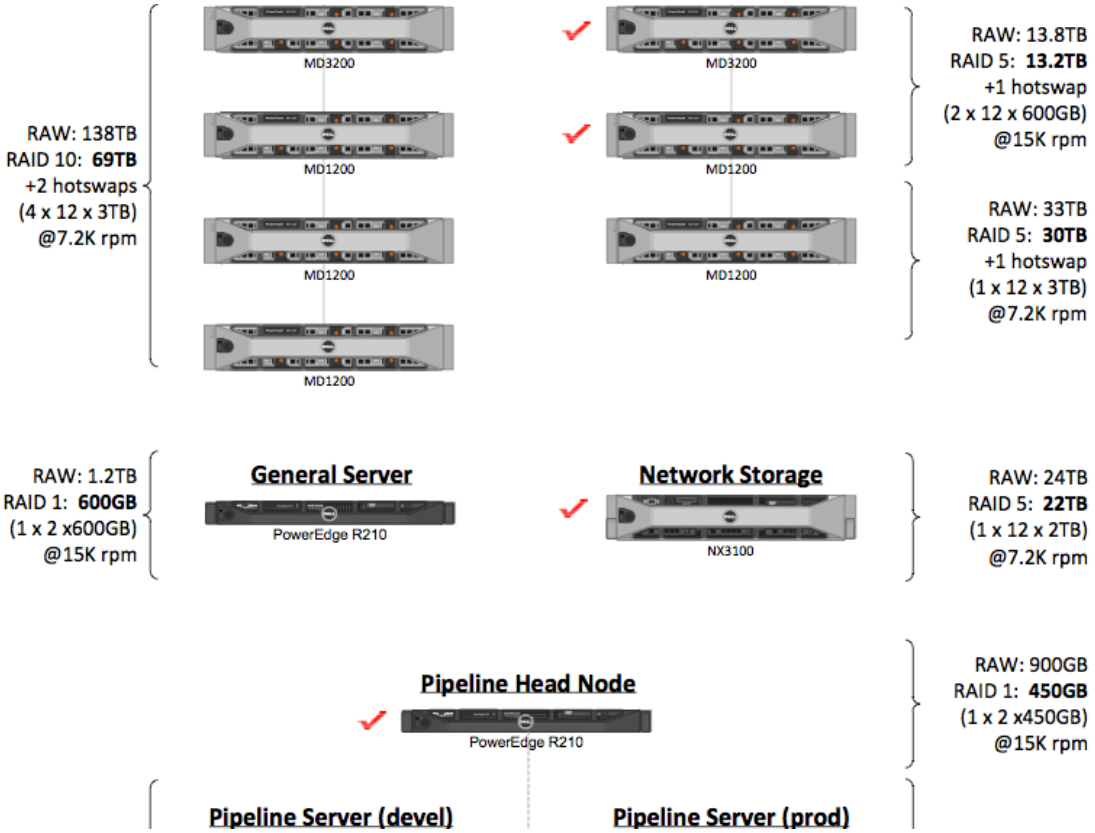
1000 Genomes Collaboration

- **Objective**:
  - Enable rapid generation of new high-confidence SVs, work with leaders in the area
- **Summary**:
  - Try to identify SVs that are most likely true positives by evaluating their genotypes, i.e. assessing Hardy-Weinberg Equilibrium (e.g., Hardy-Weinberg p-value) or other metrics based on population genotyping.
- **Method**:
  - Perform post-processing on BreakSeq calls made in several individuals in a population
  - Search for SVs that tend to yield clearcut genotypes (based on assessing reference and alternate alleles for BreakSeq across the analyzed individuals.
- **Validation**:
  - The quality of the new SV set could be assessed using existing validation data from the 1000 Genomes Project (e.g., aCGH and SAV, as approaches for assessing FDRs with these are already in place) or
  - By generating additional PCR validations.
- **Benefits:**
  - Latest junction library from the 1KG
  - Expert advice on the methodology
  - Improvement of BreakSeq
  - High-quality SV call set for filtering for accuracy
  - Begin to build an SV frequency database, which could also be used for disease association
- **Current Status:**
  - Finished running BreakSeq on 991 1KG genomes based on a deletion junction library with 2 different stringency levels.

Scientific Computing Infrastructure

- Looking for a scientific computing manager
  - Hands-on experience with HPC, Big data, Cloud computing, Data security
- Continuing to anticipate our 3-6 month computing needs
  - R&D needs are priority currently, ops needs will become more prominent later
- New hardware geared at:
  - Increasing our genome processing capability
    - Dell R910 servers x 2 = ~Four 30x genomes processed/day if fully utilized
  - Increasing our overall storage to enable development work, download of numerous genomes for development and testing work requires more storage
    - DAS: Dell MD3200 x2, MD1200 x 4 in RAID 5 and RAID 10 = 112 TB storage
    - NAS: Dell NX3100 = 22 TB Storage
  - Hosting key development software
    - Dell R210 for hosting JIRA, GIT, TWIKI, NIS
- Other items
  - Server backup planning
  - Move to new location

Scientific Computing Infrastructure



Personalis EX2042

Pipeline Performance & Capacity

			All Stages	
	Genomes	**Coverage (X)**	**Real Time (sec)**	**Performance (sec/X)**
Sequential	1	49	99,385	2,027
Parallel	3	161	200,520	1,246
Improvement	-	-	-	1.63

Running in Parallel:	**Average**
30X Coverage (hrs)	10.38
Genomes/node/day	2.31
2-node capacity	4.62

Personalis EX2042

Key R&D Risks and Mitigation

Risk	Mitigation
• Hiring great people takes time. If not hired on schedule, may impact schedule	• Recruiters now involved • Web site careers posting
• Numerous development tasks are still research in nature making timeline more difficult to estimate	• Continue rapid prototyping to resolve, get visibility on some key technical risks to plan appropriately, especially with Accuracy effort, GIS product and report generation
• Long lead time content curation and content quality items for GIS and RVDS	• Already started on those items with long lead times to develop concurrently with Pipeline (MendelDB, etc)
• Complexity of development on multiple products fronts with small development team	• Hiring • Basic Project Mgmt infrastructure being deployed including JIRA, GIT, TWIKI, etc to enhance efficiency
• Licensing of applications and content for analysis	• Created process for systematic review and acquisition of licenses. Prioritized most important ones (over 50 on the list!)
• Ethical use of Genomes for R&D	• Ethics board approval for all Genomes used for R&D • Part of orientation training
• Product Differentiation	• Continuing to push differentiation in three key areas: Accuracy/Quality, Content, Features
• Scalability of products and services	• Approach is to prioritizing resources towards demonstrating value with our services, then scaling once that is demonstrated
• Identifying high quality collaborators for testing our services. Potentially long lead time.	• Begin to identify collaborators for each of the services to get IRB ethical legal framework in place

Summary of R&D Next Steps & Priorities

- Continuing hiring for open positions
- Continue development on Pipeline Service with Aug 2012 prototype target
- Begin to identify friendly collaborators for testing Pipeline Service Prototype
- Further refine specs and project plan around Rare Variant Discovery Service
- Develop functional specification and detailed project plan for Genome Interpretation Service
- Continue to execute on content priorities
- Continue to execute on Accuracy objectives and integration with laboratory capabilities