## **Using Galaxy for Molecular Assay Design**

2012 Galaxy Community Conference

J Ireland, Principal Bioinformatics Scientist

#### **Galaxy at 5AM Solutions**

- 5AM: Custom software and bioinformatics services company
- Galaxy converts for one year
  - http://info.5amsolutions.com/blog/bid/86705/DIY-Bioinformatics-A-Whole-New-Galaxy
- Galaxy in action
  - Internal bioinformatics resource
  - Client 1: data source integration
  - Client 2: custom workflows for assay design
  - Client 3 (in progress): deploy Galaxy for assay design



## What is Molecular Assay Design (MAD)?

#### Design

 In-silico process to identify oligos (probes, primers) to detect and/or quantify nucleic acid targets

#### Balance

- Target Coverage: oligo location, expected performance
- Cost: \$\$\$, oligo real-estate, oligo interactions

#### Technologies

o PCR, TaqMan, molecular inversion probes, LNA, oligo arrays, ...

#### Applications

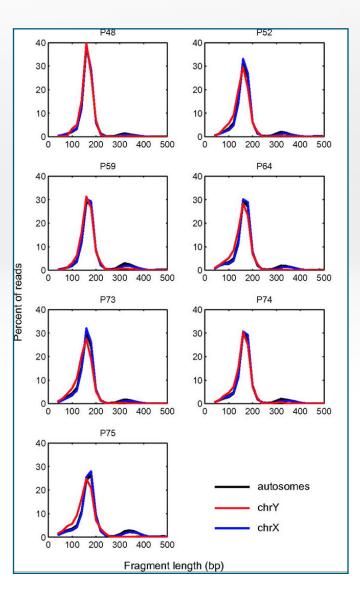
o Targeted sequencing, genotyping, gene expression, miRNA, ...

#### The MAD Workflow

- Target Selection
  - Identify locations on genome/transcriptome of interest while avoiding trouble regions
- Assay Design
  - Select sets of primers and/or probes against targets
- Assay Prioritization/Filter
  - Filter, prioritize and annotate assay designs
- Ensemble Selection
  - Select final combination(s) of assays

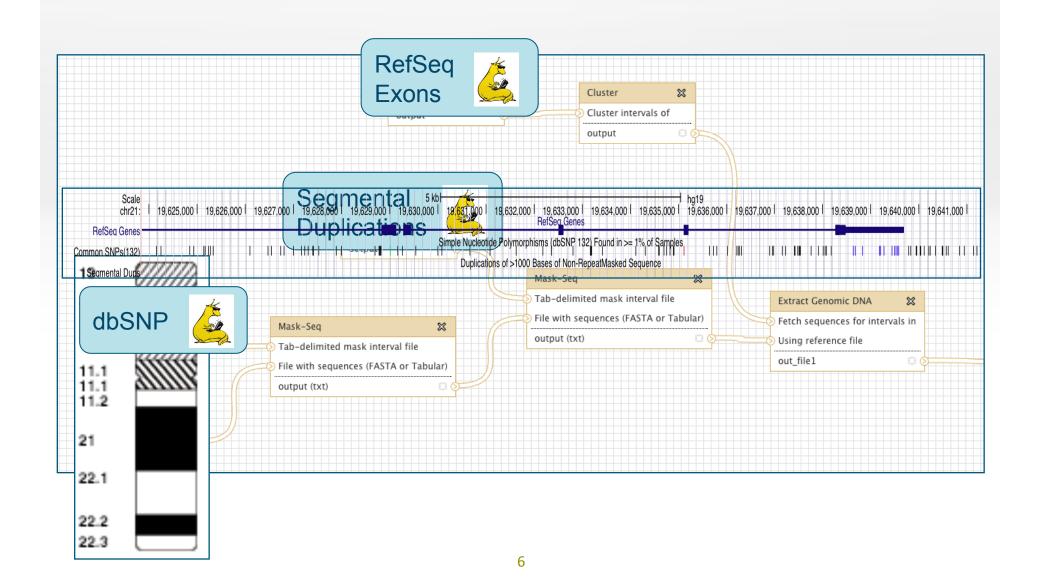
#### **Cell-Free DNA Testing – A MAD Example**

- Client developing a multiplexed PCR assay for detection of trisomies in cfDNA
- Requires sensitive assay to detect fetal cfDNA in high background of maternal cfDNA
- Short assay lengths needed because cfDNA highly fragmented
- Genetic variation (SNPs, CNV) can confound detection

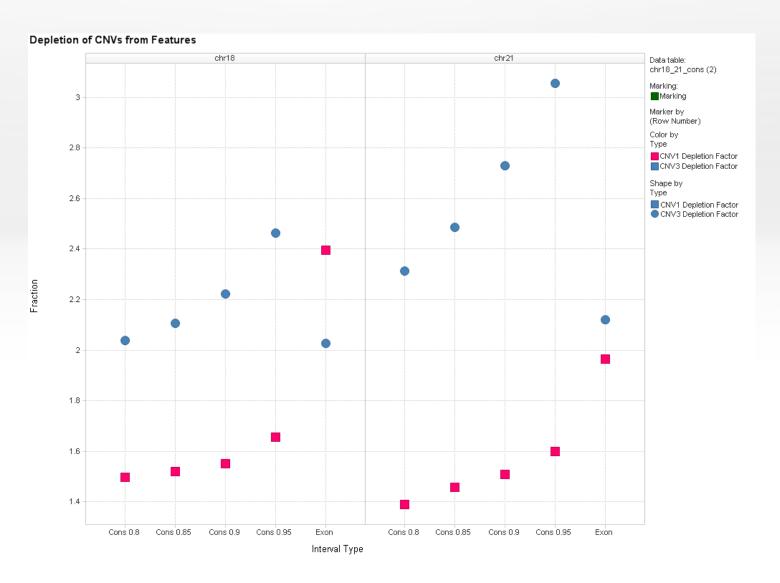


Fan, et al. Clin. Chem, 2010

#### **Targets for cfDNA Testing**



# An Aside: Galaxy Powered MAD Analysis



## **Assay Design**

	_		~							
#AssayId	For Primer	Rev Primer	Amp Size Amp Seq	Comp	l End Compl Any	Target Id	Chrom	Amp Start	Amp Stop   F	or Primer TM
chr21:9907188-9908432_PR1	TAAATTACGAGGGTCCTCCC	GGAATGGAAAGAACTGAGCC	121 TAAATTACGAGGGTCCTC	CCCAGGGAGCGCGGGCC	0 3	chr21:9907188-9908432	chr21	9907620	9907741	59.793
chr21:9907188-9908432_PR2	GCATCCGTTAACAAGGGAAA	CCACCTCTGGTGGAAAGTAT	82 GCATCCGTTAACAAGGGA	AAACAAAAGGACGGCCC	1 3	chr21:9907188-9908432	chr21	9907832	9907914	60.018
chr21:9907188-9908432_PR3	ACAAGGGAAACAAAAGGACG	CCACCTCTGGTGGAAAGTAT	72 ACAAGGGAAACAAAAGG	ACGGCCCATCTATGCC1	0 4	chr21:9907188-9908432	chr21	9907842	9907914	60.162
chr21:9907188-9908432_PR4	GCAGGGAAGGAAGAAGGTA	TTTCCCTTGTTAACGGATGC	64 GCAGGGAAGGAAGAAGG	GTACTTAGGCTCAGAAG	1 6	chr21:9907188-9908432	chr21	9907788	9907852	60.604
chr21 <del>-9907188-9908432_PR5</del>	GGCTCAGTTCTTTCCATTCC	TACCTTCTTCCTTCCCTGC	87 GGCTCAGTTCTTTCCAT	CCTGATATTTGACGGA	0 3	chr21,9907188-9908432	chr21	9907721	9907808	60.09
chr21:9907188-99089032 PR6	TCGATGAAGAGACTGTTGGT	CCTCAAGGACATCCACATGA	50 bases TCGATGAAGAGACTGTTC	GTCATGGCAGTGACGT		chr21:9907188-9908432	chr21	9907313	9907379	59.943
chr21:9907188-9908422: PR7		32:500 grt9:632:510 rg19:632.5				<b>9</b> .682.599907198832960084319.			9.692.83009	60.018
chr21:9907188-9908432 1788 U	C FRITTEN TWITTEN TOWN	CHALLER FLATE HALPENDAG	AFTAATCC/AAQAGCCCCTGTAGAAAAGC	CITATICITACAAA ICAACCAG			TANT ZAC	Ced (9) 83 2	19997969	AAGAACI <sub>60.018</sub>
chr21:9907188-9908432_PR9	GGTACTTAGGCTCAGAAGGG	hr21:196325T6-19632603 PA4	110 CCTACOTTAGGCTCAGAA	User Supplied Track	3	chr21-0007188-0008432	chr21	9907804	9907914	60.454
chr21:9907188-9908432_PR10	GCAGGGAAGGAAGGTA	hr21-19632516-19632603TBB1	AGGA AGAR			88-9908432	chr21	9907788	9907862	60.604
chr21:9909046-9909277_PR1		hr21-19632518-19632603-PRT0	CAGTCC			09046-9909277	chr21.	9909159	9909246	60.533
chr21:9909046-9909277_PR2		chr21419632516419632603APR8	A DO	CACCCACACCTIC			chr21.	9909160	9909246	60.534
chr21 9909046-9909277 PR3		chr219632516-19632603 PR7	SWETGE!			46-9909277	chr21.	9909159	9909247	60.533
chr21:9909046-9909277_PR4	AAGGATGCAGTCCACCAG	7chr24:19632516-19632603APR	) regad			09046-9909277	chr21.	9909160	9909247	60.534
chr21 9909046-9909277 PR5	GAAGGATGCAGTCCACCAG	Achr24/49632546-49632603APR	6 reenes	TODA COCHO AGOSTE			chr21	9909159	9909247	61.66
chr21:9909046-9909277_PR6	AGGATGCAGTCCACCAG	chr21:19632516-49632603_PR				46-9909277	chr21	9909161	9909246	59.765
chr21 9909046-9909277 PR7	GAGGTGTCGCGAAGGAT	CCAAGT: Chr21:19632516:1963	2603_PR5 <del></del>	JCAGTOCKCEAGUA			chr21	9909149	9909246	60.277
chr21 9909046-9909277 PR8	GAAGGATGCAGTCCACCA	ACCAAG: 01/2E1963254671963				46-9909277	chr21	9909159	9909247	60.533
chr21 9909046-9909277 PR9	AAGGATGCAGTCCACCAG	ACCAAGTTCACACAGGACA	87 AAGGATGTAGTCCACCAC	CHelSed Generalics	0 4	chr21-9909046-9909277	chr21	9909160	9909247	60.534
chr21 9909046 9909279 PR10	GAAGGATGCAGTCCACCA	AAGTTCACACAGGACAATCT	OD GARGGAT BOARTE CACCA	IDEACGACABGILISE	40/ <sup>2</sup> -k O	CI112.1.9909040-9909277	chr21	9909159	9909244	60.533
chr21 9915249-9916547 PR1	CCCGCTTCTCAAAACTCATT	TGACACTCCGGGCATATAAA	111 The Profestions of \$100	DISTIS (005) VP 132) Found III	Sociones 4	chr21:9915249-9916547	chr21.	9916317	9916428	60.018
chr21 9915249 9916547 PR2	CCCGCTTCTCAAAACTCATT	ATGACACTCCGGGCATATAA	111 CCCGCTTCAAAACTCA	ATTGCACAATAGGTTGA	4	chr21:9915249-9916547	chr21.	99:1631.7	9916429	60.018
chr21:9915249-9916547_PR3	GAGGGACTTCTCTACACCAG	TGACACTCCGGGCATATAAA	79 GAGGGACTTCTCTACACC			chr21:9915249-9916547	chr21	9916349	9916428	60.089
chr21:9915249-9916547_PR4	AAAACATACAGCTTACCCGC	TGACACTCCGGGCATATAAA	126 AAAACATACAGCTTACCO	CGCTTCTCAAAACTCAT	1 4	chr21:9915249-9916547	chr21	9916302	9916428	60.091
chr21:9915249-9916547_PR5	CCCGCTTCTCAAAACTCATT	GCCTACAACAGACACATCAG	132 CCCGCTTCTCAAAACTC	ATTGCACAATAGGTTGA	2 3	chr21:9915249-9916547	chr21	9916317	9916449	60.018
chr21:9915249-9916547_PR6	CCCGCTTCTCAAAACTCATT	CTGCCTACAACAGACACATC	134 CCCGCTTCTCAAAACTC	ATTGCACAATAGGTTGA	0 3	chr21:9915249-9916547	chr21	9916317	9916451	60.018
chr21:9915249-9916547_PR7	GAGGGACTTCTCTACACCAG	ATGACACTCCGGGCATATAA	80 GAGGGACTTCTCTACACC	CAGGACCTATTATTAAA	2 3	chr21:9915249-9916547	chr21	9916349	9916429	60.089
chr21:9915249-9916547_PR8	AAAACATACAGCTTACCCGC	ATGACACTCCGGGCATATAA	127 AAAACATACAGCTTACCO	CGCTTCTCAAAACTCAT	2 4	chr21:9915249-9916547	chr21	9916302	9916429	60.091
chr21:9915249-9916547_PR9	GAGGGACTTCTCTACACCAG	GCCTACAACAGACACATCAG	100 GAGGGACTTCTCTACACC	CAGGACCTATTATTAAA	1 3	chr21:9915249-9916547	chr21	9916349	9916449	60.089
chr21:9915249-9916547_PR10	GAGGGACTTCTCTACACCAG	CTGCCTACAACAGACACATC	102 GAGGGACTTCTCTACACC	CAGGACCTATTATTAAA	1 3	chr21:9915249-9916547	chr21	9916349	9916451	60.089
chr21:9968515-9968593_PR1	ACAGGGACATAGAACCAAGC	ATTTTCCTCCGGAAGTGCG	76 ACAGGGACATAGAACCA	AGCCCCAGGCCTGCTCA	1 4	chr21:9968515-9968593	chr21	9968517	9968593	60.967
chr21:9968515-9968593_PR2	ACAGGGACATAGAACCAAGC	TTTTCCTCCGGAAGTGCG	75 ACAGGGACATAGAACCA	AGCCCCAGGCCTGCTCA	1 3	chr21:9968515-9968593	chr21	9968517	9968592	60.967
chr21:9968515-9968593_PR3	GACATAGAACCAAGCCCCAG	ATTTTCCTCCGGAAGTGCG	71 GACATAGAACCAAGCCC	CAGGCCTGCTCAGCTAC	1 3	chr21:9968515-9968593	chr21	9968522	9968593	61.685
chr21:9968515-9968593_PR4	CAGGGACATAGAACCAAGCC	ATTTTCCTCCGGAAGTGCG	75 CAGGGACATAGAACCAAG	GCCCCAGGCCTGCTCAC	1 4	chr21:9968515-9968593	chr21	9968518	9968593	61.685
PRIMER INTERNAL MAX	SIZE=36	# Max probe lengt	n	<u>'</u>						
PRIMER INTERNAL OPT		# Opt probe lengt						]		
PRIMER INTERNAL MIN		# Min probe TM (S								
PRIMER INTERNAL OPT		# Opt probe TM (S								
DOTUME THE PROPERTY AND		# Ven probe IN (S								

# No G at 5' end

PRIMER INTERNAL MAX TM=80.0

PRIMER\_MAX\_SELF\_ANY=8.0

INTERNAL 5PRIME BASE EXCLUDE=G

PRIMER PAIR MAX COMPL ANY=8.0

PRIMER\_INTERNAL\_MAX\_SELF\_ANY=8.0

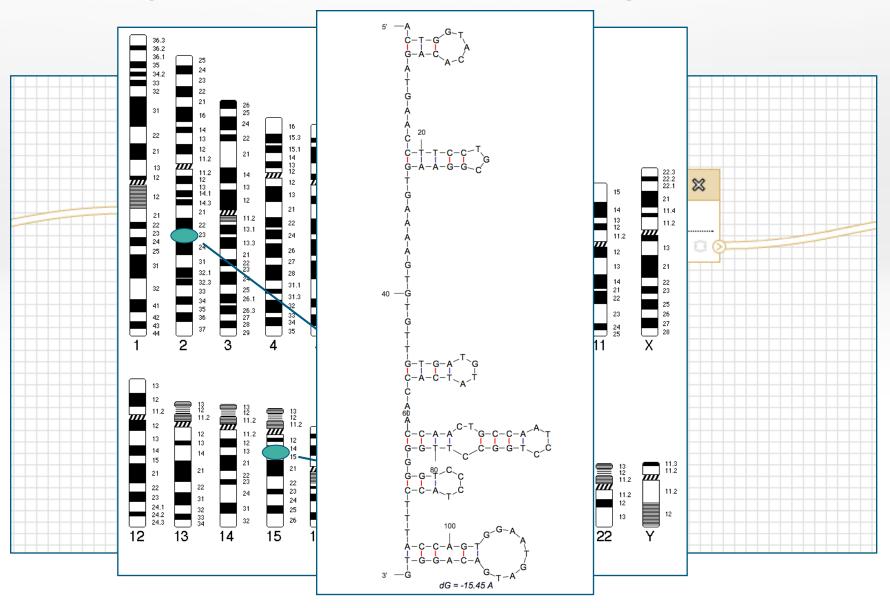
# Max probe TM (Santa Lucia)

# Pair compl (anywhere)

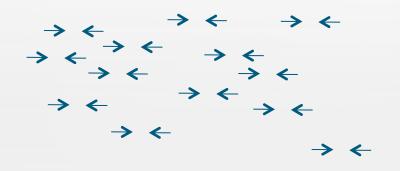
# Primer annealing (anywhere) to itself

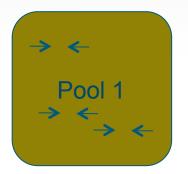
# Probe annealing (anywhere) to itself

## **Assay Prioritization and Filtering**



#### **Assay Ensembles**







- Combine multiplexed pools
  - Avoid misamplification
  - Avoid oligo interactions
- Select "optimal" subset
  - Coverage
  - Expected performance
  - Best multiplex
- Programs to help
  - o MultiPLX

# **Looking Back**



No more Perl soup

 Data sets, parameters, steps preserved

Reproducible design

Sharable with client

## **Looking Ahead**

Tool Shed submission

Extend workflows to ensemble selection



- Automate report generation
  - Sweave for Galaxy?

#### Thanks!



**Andy Evans** 



Will FitzHugh

The Galaxy Community!