Complex diseases
A mouthful of Scope®
The power of the Illumina platforms

Annette T. Lee, PhD
The Robert S. Boas Center for
Genomics and Human Genetics
The Feinstein Institute for Medical Research
North Shore-LIJ Health System

GWAS Groups

- Autoimmune Diseases
  - iSelect Custom Beadchip
- Diabetic Nephropathy
  - BeadXpress GoldenGate
- Chronic Lymphocytic Leukemia
  - Copy Number Variation
  - Methylation
  - Gene expression
  - Micro RNA expression

Rheumatoid Arthritis and Systemic Lupus Erythematosus

- Inflammatory autoimmune disease
- Clinically heterogeneous
- Presence of disease specific antibodies
- Affects females > males
- Complex mode of inheritance
- Multifactorial – genetic and environmental

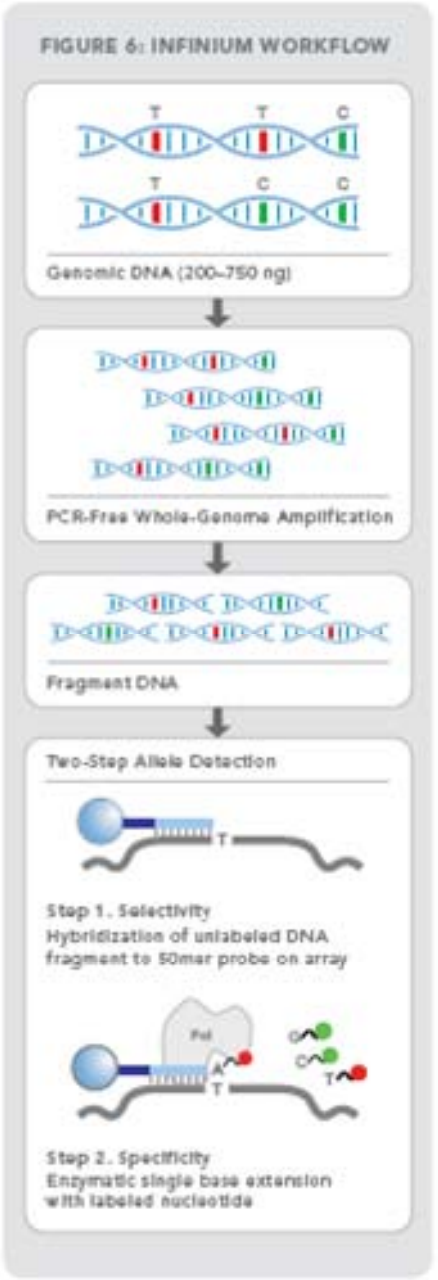
Rheumatoid Arthritis and Systemic Lupus Erythematosus

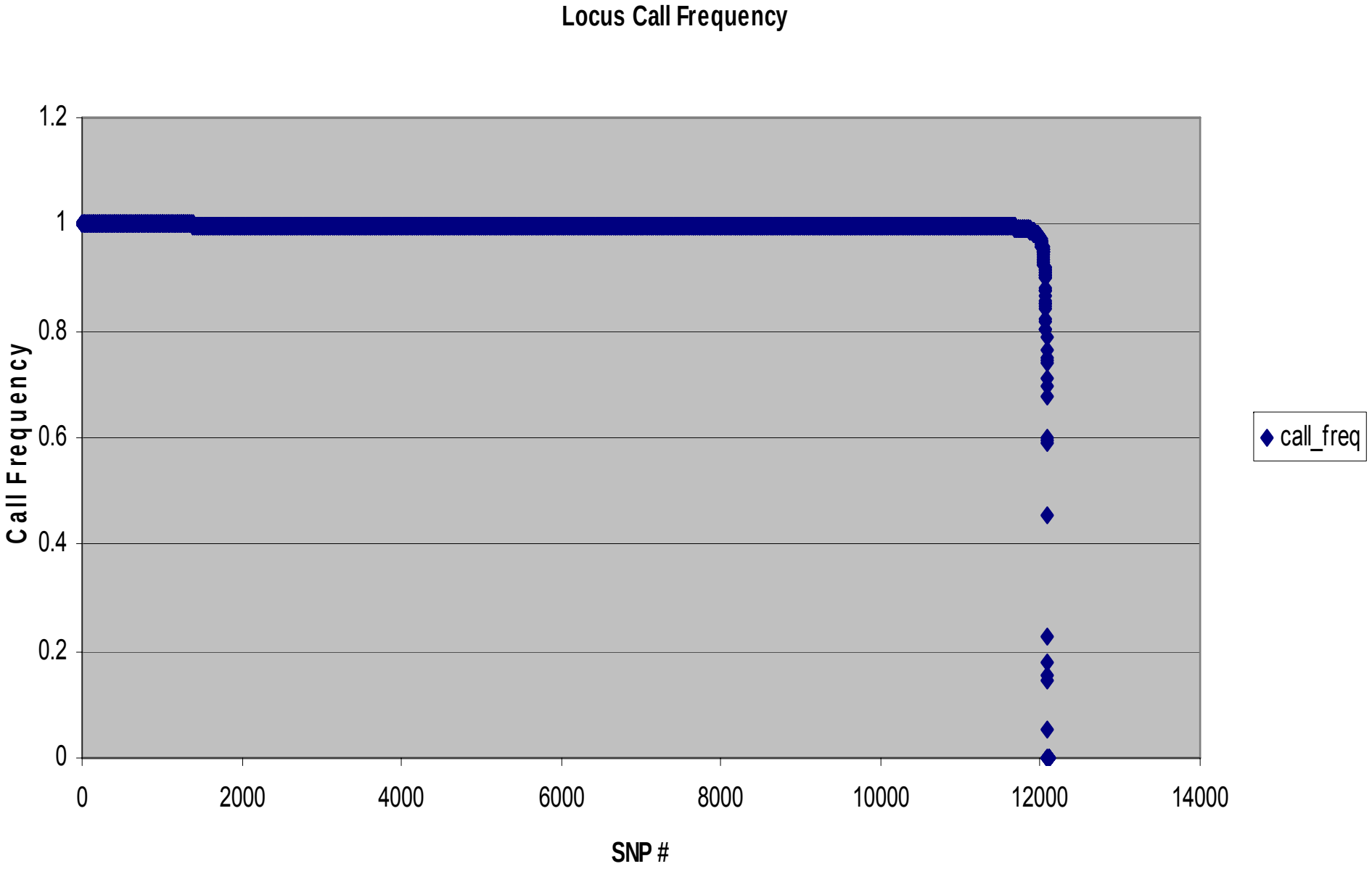
- Genetic components
  - Familial Clustering
  - Linkage studies
  - Candidate genes (PTPN22, TNFAIP3, STAT4)
- GWAS
  - 1522 RA cases and 1850 controls
    - TRAF1-C5 (Plenge et al 2007 357:1199)
  - 1311 SLE cases and 1783 controls
    - C8orf13-BLK, ITGAM-ITGAX (Hom et al NEJM 2008 358:900)

iSelect Custom Infinium Beadchip

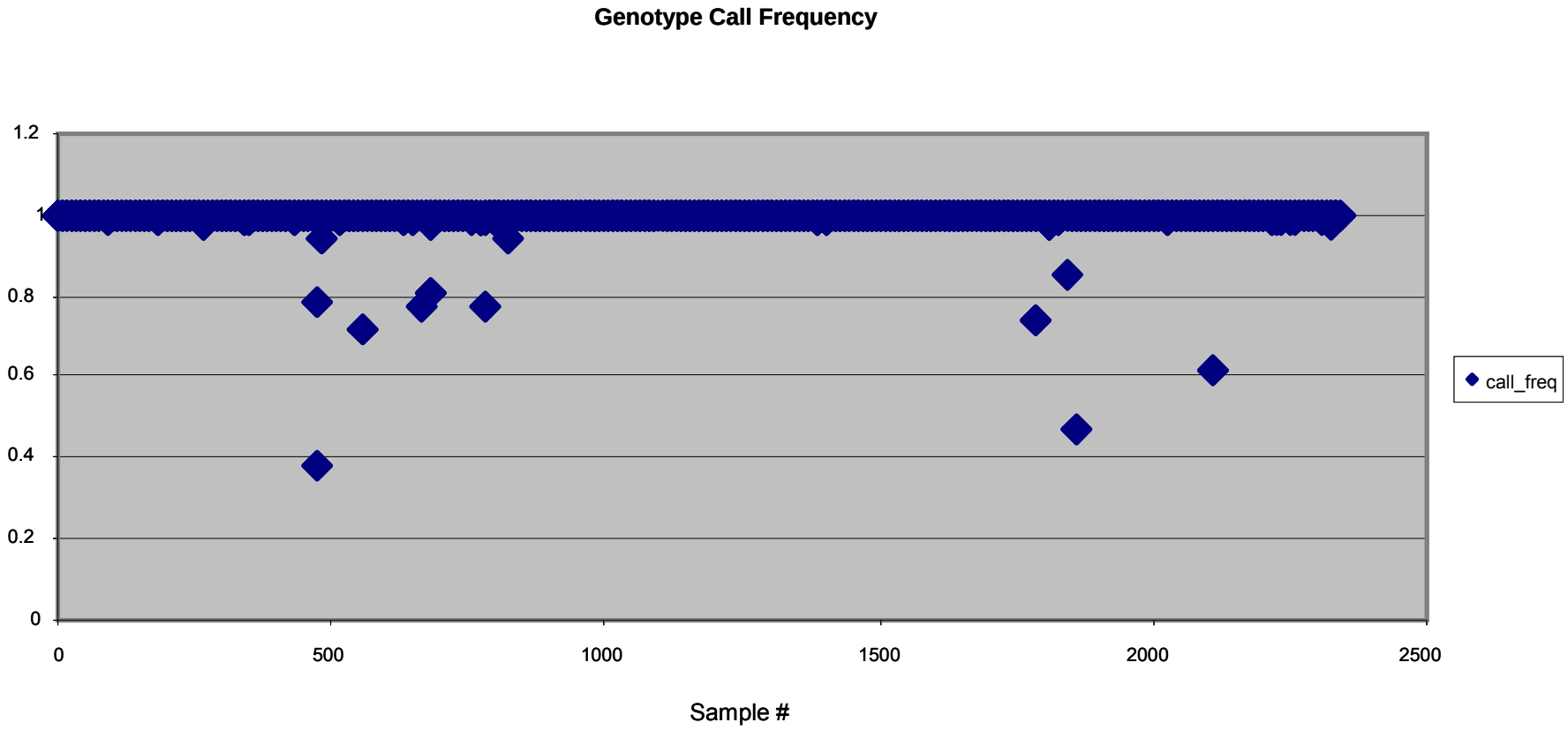
- 13,680 beadtypes
  - 12048 1 beadtype + 816 2 beadtype = 12,864 SNPs
    - ~6000 Ancestry informative markers
    - ~6000 Autoimmune disease specific SNPs
  - 8227 SNPs GWAS results + SNPs surrounding risk loci, reported SNPs and rare variants

- Replication set
  - ~3000 SLE
  - ~2000 RA
  - ~2000 other autoimmune diseases
  - ~1000 controls





Locus success rate 99.4% (12037/12107)



RA Sample success rate >95%
99.9% (2281/2293)

Average call frequency 99.5% (95.9-99.81%)

Duplicate Reproducibility

Rep1_DNA_Name	Rep2_DNA_Name	# Correct	# Errors	Total	Repro_Freq
00AD4073_4305901300_G	00AD4073_4309241089_B	12070	1	12071	0.999959
02AD4342_4305901349_H	02AD4342_4309241089_F	12068	2	12070	0.999917
03AD4540_4309241006_G	03AD4540_4309241021_B	12074	2	12076	0.999917
03AD4878_4305901232_G	03AD4878_4309241095_G	12067	0	12067	1
03AD5065_4305901231_A	03AD5065_4305901233_F	12078	0	12078	1
03AD5189_4305901301_A	03AD5189_4309241095_D	11985	0	11985	1
04AD5441_4309241021_F	04AD5441_4309241034_H	12071	1	12072	0.999959
04AD6000_4305901239_G	04AD6000_4305901316_G	12070	0	12070	1
04AD6239_4305901219_A	04AD6239_4305901316_D	12062	0	12062	1
04AD7123_4305901426_F	04AD7123_4305901426_J	12027	2	12029	0.999917
05AD8031_4305901317_D	05AD8031_4305901317_G	12070	0	12070	1
05AD8514_4305901240_K	05AD8514_4305901469_D	12073	0	12073	1
05AD8717_4305901466_F	05AD8717_4305901466_J	12076	0	12076	1
05AD8859_4305901241_F	05AD8859_4305901469_B	12072	0	12072	1
08AD7471_4305901032_F	08AD7471_4305901343_A	12059	2	12061	0.999917
91M0194_4305901257_D	91M0194_4305901257_G	12076	0	12076	1
92M0768_4305901268_D	92M0768_4305901457_K	12074	0	12074	1
93M0265_4305901273_F	93M0265_4305901404_B	12031	2	12033	0.999917
94M1034_4305901268_G	94M1034_4305901268_J	12073	0	12073	1
94M1339_4305901321_B	94M1339_4305901348_G	12067	0	12067	1
97AD1665_4305901283_H	97AD1665_4305901321_F	11892	1	11893	0.999958

Reproducibility 100% (21/21)

Beadchip Concordance

- 2 samples
  - both whole genome and iSelect chip

KNZ725263	8223/8227	99.95%
KNZ530311	8226/8227	99.99%

Manuscript submitted

Type 1 Diabetic Nephropathy

- Persistent Proteinuria
  - Incidence of 30-40%
- Leading cause of end stage renal disease
- Familial clustering
- Multifactorial –
  - Genetic and environmental factors

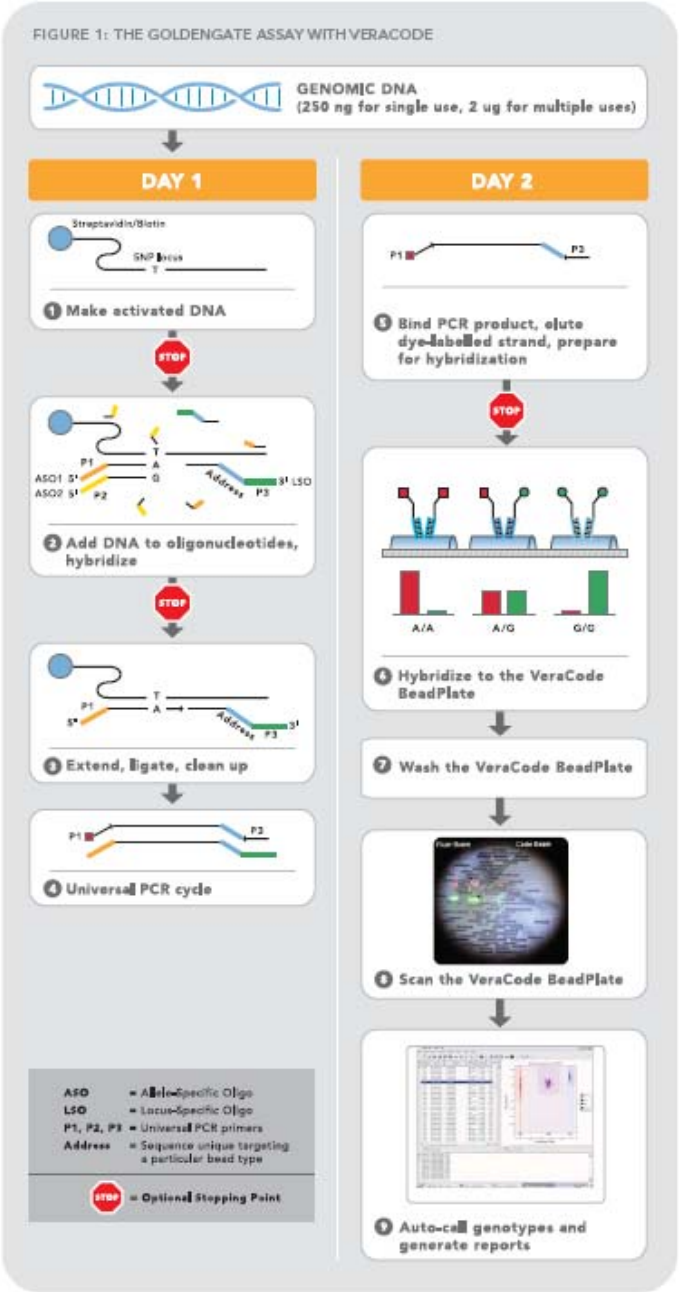
Type 1 Diabetic Nephropathy

- Genetic component
  - Familial clustering
  - Linkage studies (3q, 4p, 9q, 16q, 22p)
  - Candidate genes (CNDP1, ELMO1, NRP1)

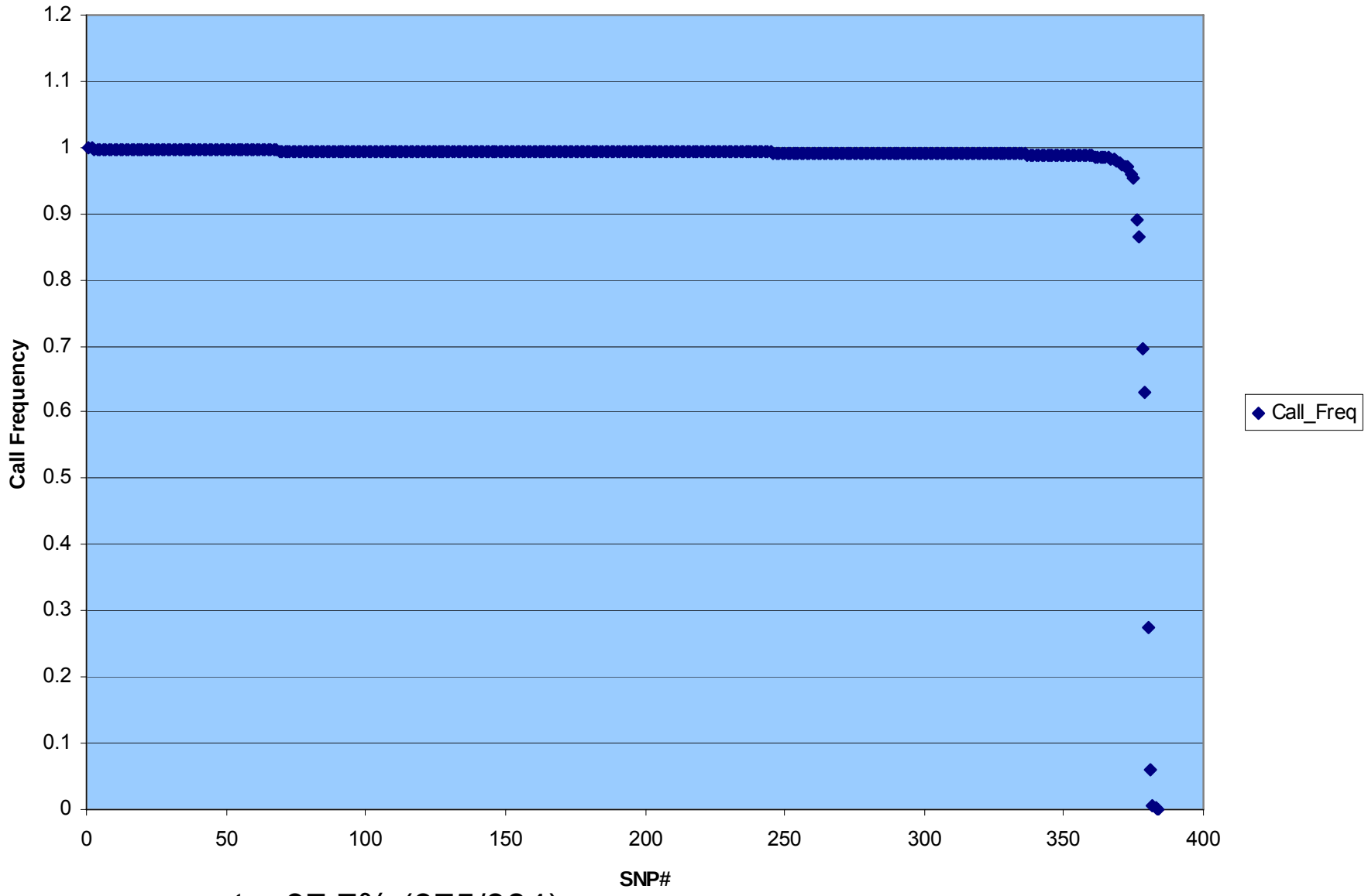
- GWAS HumanHap300 v1
  - 188 Type 1 diabetic subjects
  - 196 Type 1 diabetic subjects with renal insufficiency

Veracode-GoldenGate Genotyping

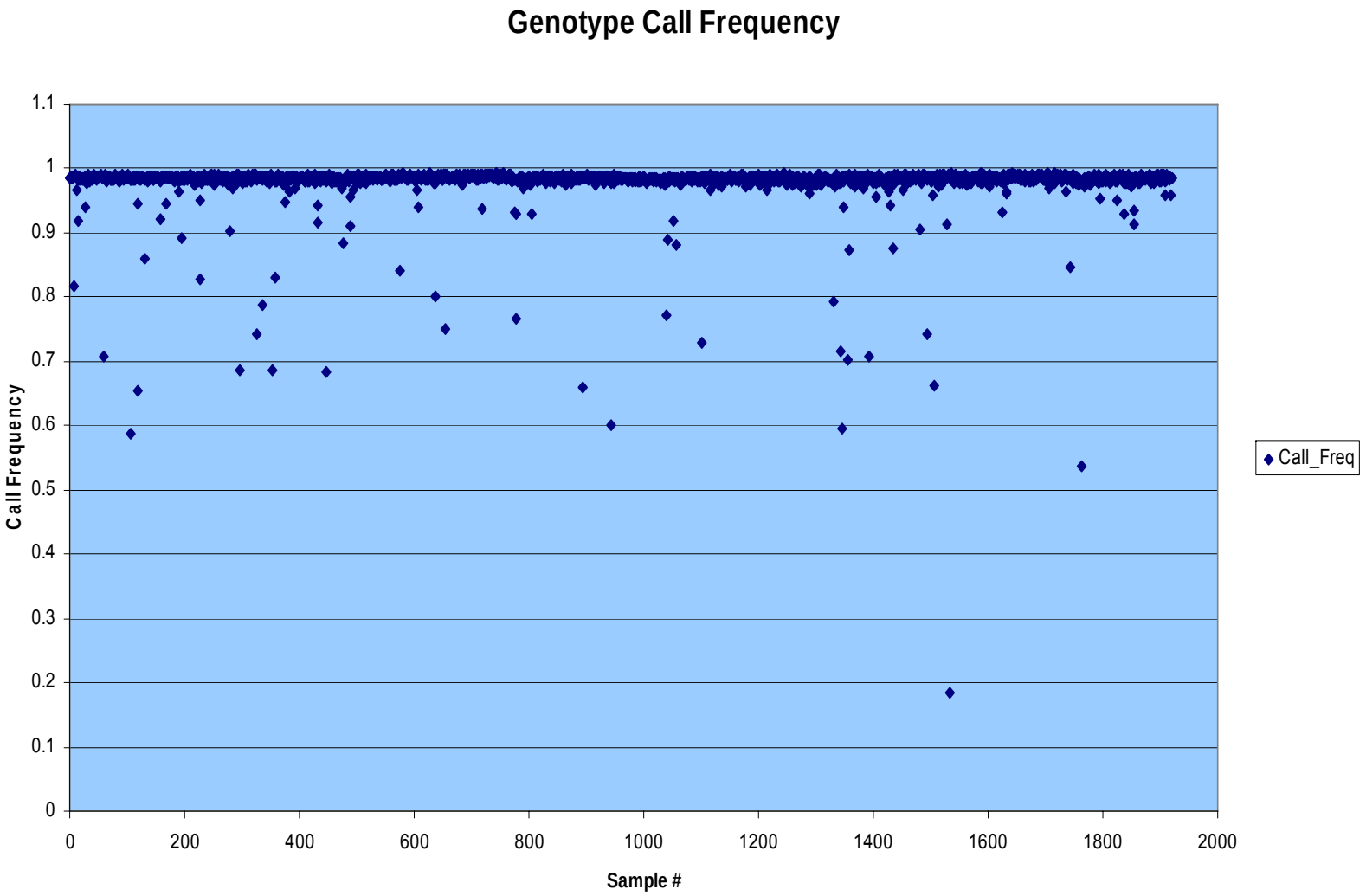
- 384 plex
  - 288 SNPs from HumanHap300 v1
  - 96 additional SNPs
    - candidate genes, reported SNPs, inferred SNPs
- Replication set:
  - 1920 samples
    - T1D normal renal function
    - T1D with renal insufficiency



Locus Call Frequency



Locus success rate 97.7% (375/384)



Sample success rate (>90%)
98.2% (1886/1920)

Average Call Frequency 98.3% (90.4 – 99.22%)

Reproducibility

Rep1	Rep2	# Correct	# Errors	Total	Repro_Freq
0012-4	0012-4C	378	0	378	1
0016-3	0016-3C	377	0	377	1
0309-3	0309-3C	377	0	377	1
0347-3	0347-3C	378	0	378	1
0532-3	0532-3C	377	0	377	1
0401-3	0401-3C	375	0	375	1
0080-3	0080-3C	373	0	373	1
0081-3	0081-3C	377	0	377	1
5274-3	5274-3C	376	0	376	1
5276-3	5276-3C	376	0	376	1
RN2910	RN2910C	371	0	371	1
RN3006	RN3006C	375	0	375	1
0109-3	0109-3C	376	0	376	1
0110-3	0110-3C	377	0	377	1
1379-3	1379-3C	374	0	374	1
1414-3	1414-3C	378	0	378	1
1909-3	1909-3C	376	0	376	1
1913-3	1913-3C	377	0	377	1

Reproducibility 100% (18/18)

Infinium to Veracode Cross Platform Concordance

- 2 samples both platforms
- 288 SNPs in common
  - 5238-3 288/288 100%
  - K0895-3 287/288 99.65%

384plex for another dataset of ~2500 samples

- Analysis in progress

Chronic Lymphocytic Leukemia (CLL)

- Elevated WBC, detected with routine lab work
- Clonal proliferation and accumulation of neoplastic lymphocytes (mostly B-cells)
- Accounts for 30% of all diagnosed leukemia cases in western countries
- Lifetime risk of developing CLL is 0.37%

Chronic Lymphocytic Leukemia (CLL)

- Rarely diagnosed under the age of 40, incidence increase with age.
  - Mean age of diagnosis = 65yrs.
- Men:women 2:1
- There is no cure. In 2007 approximately 15,500 new cases were diagnosed and 4,500 deaths were reported.

Prognostic Characteristics

	Aggressive	**Benign**
Clinical course		
IgVH gene mutations	Few or none	Significant numbers
CD38 expression	High	Low
ZAP-70 expression	High	Low
Telomere length	Uniformly short	Variable lengths
Telomerase activity	High	Low
Chromosomal Abnormalities	Δ11q22-q23, Δ17p13 trisomy 12	Δ13q14
Treatment requirement	Yes	No
Median survival (years)	6-8	> 25

Evidence for Genetic Components

- Multiple-case families
  - (vertical transmission)
- Familial Clustering
- Elevated familial risk in first degree relatives
  - (3-7 fold)
- Anticipation
  - (decreased age of onset ~20yrs)
- Variable incidence in different populations
  - (elevated Western Europe and North America)

CLL –Genetic Studies

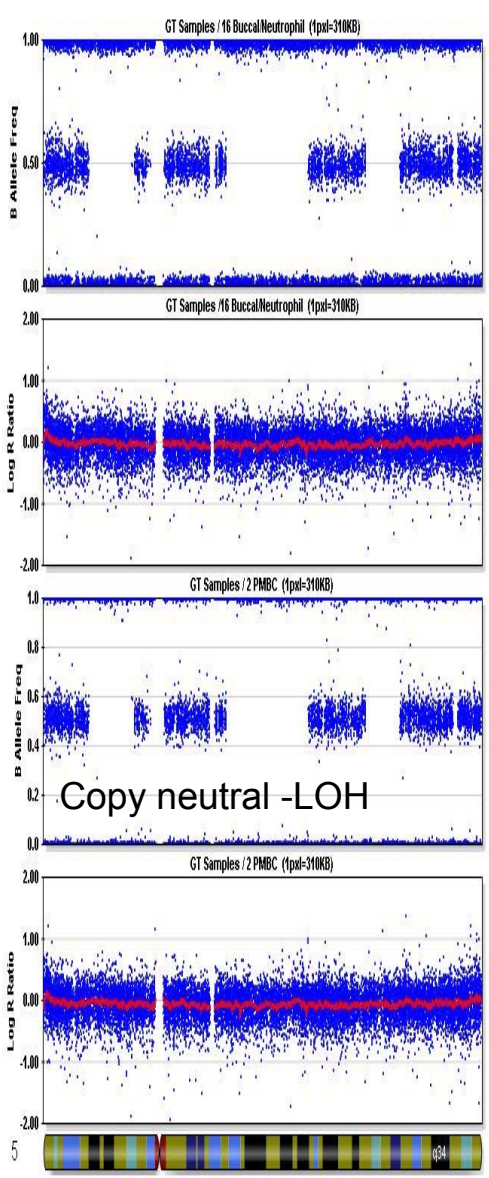
- Linkage studies
- Candidate gene studies
- Genome-wide association studies

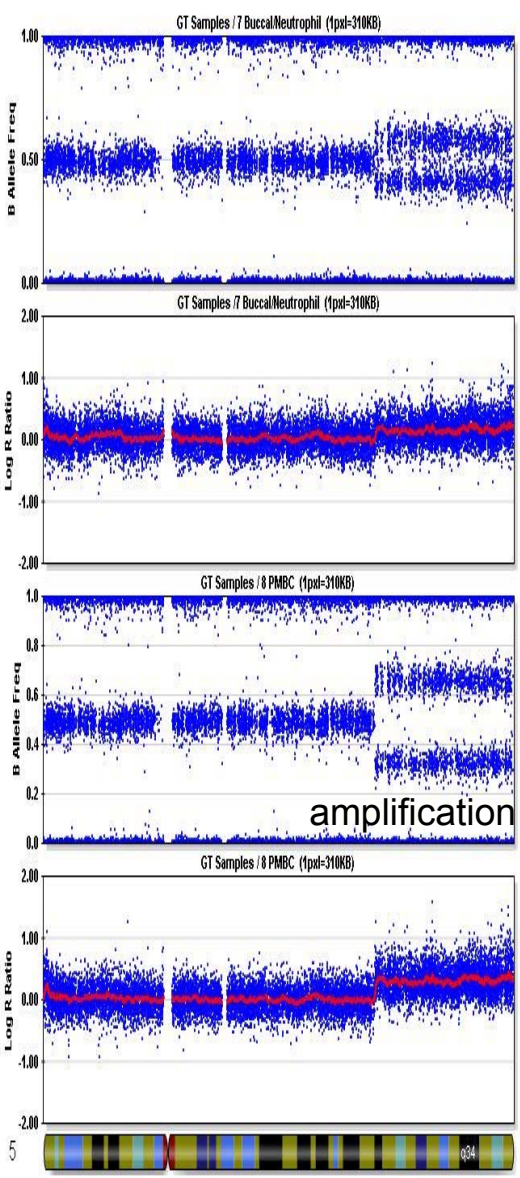
Beyond Classical Genetics

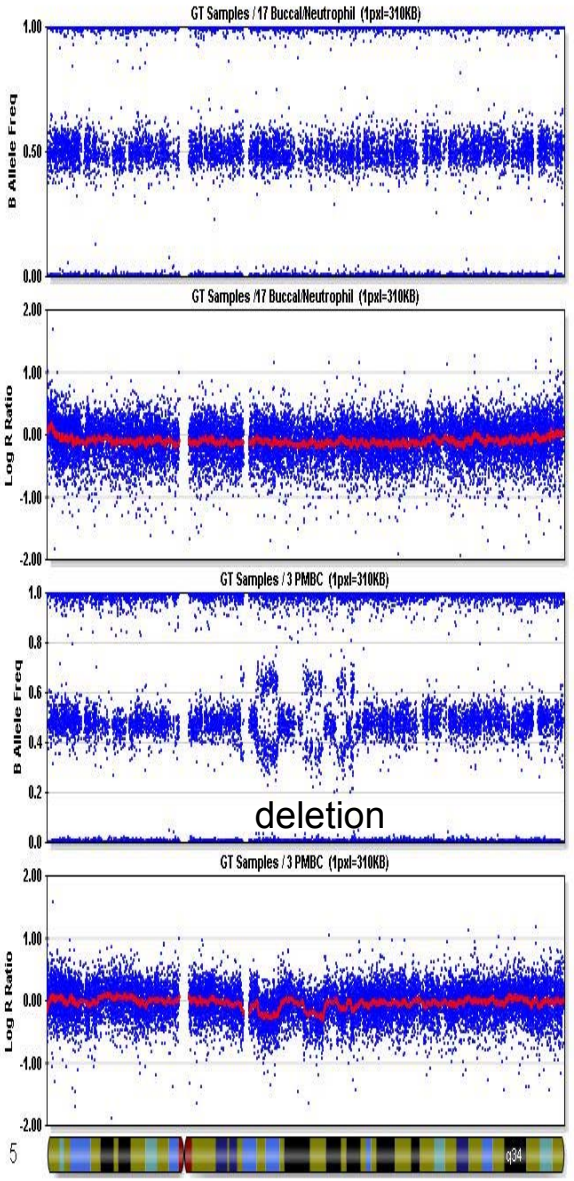
- Copy Number Variation
  - Genome-wide analysis
- DNA Methylation
  - Targeted assays
  - Genome-wide assay
- Gene Expression
  - Genome-wide assay
- Micro RNA Expression
  - Genome-wide assay

Copy Number Variation (CNV)

- Epigenetic
  - De novo or inherited
- Increase/decrease gene dosage
  - Affect gene expression levels
  - No effect


GT Samples / 16 Buccal/Neutrophil (1pxl=310KB)
B Allele Freq
GT Samples /16 Buccal/Neutrophil (1pxl=310KB)
Log R Ratio
GT Samples / 2 PMBC (1pxl=310KB)
B Allele Freq
Copy neutral -LOH
GT Samples / 2 PMBC (1pxl=310KB)
Log R Ratio


GT Samples / 7 Buccal/Neutrophil (1pxl=310KB)
B Allele Freq
GT Samples /7 Buccal/Neutrophil (1pxl=310KB)
Log R Ratio
GT Samples / 8 PMBC (1pxl=310KB)
B Allele Freq
amplification
GT Samples / 8 PMBC (1pxl=310KB)
Log R Ratio


GT Samples / 17 Buccal/Neutrophil (1pxl=310KB)
B Allele Freq
GT Samples /17 Buccal/Neutrophil (1pxl=310KB)
Log R Ratio
GT Samples / 3 PMBC (1pxl=310KB)
B Allele Freq
deletion
GT Samples / 3 PMBC (1pxl=310KB)
Log R Ratio

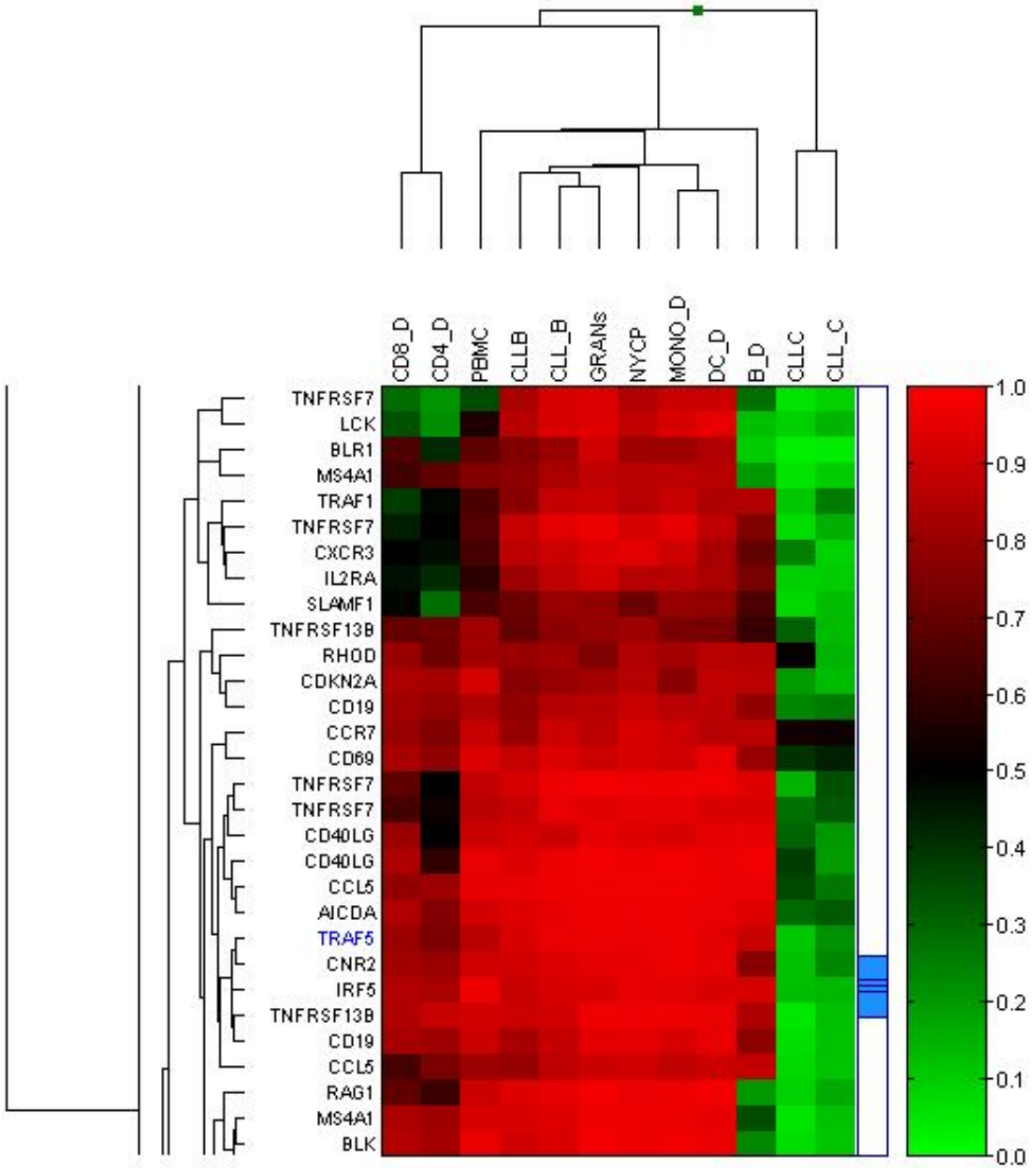
DNA Methylation

- DNA Methylation
  - Epigenetic mechanism of gene regulation
    - Heritable (X inactivation, genomic imprinting)
    - Dynamic (development, differentiation)
  - Constitutive genes: unmethylated
  - Regulated genes: methylated or unmethylated

DNA Methylation and CLL

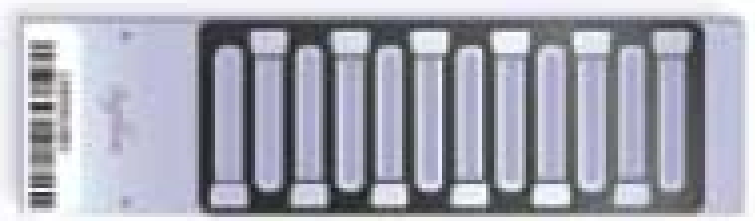
- PBMC CLL DNA is globally hypomethylated
  - Tumor suppressor genes are hypermethylated
- Some genes differentially methylated in CLL
  - ZAP-70
  - TRAF1
  - DAPK1
  - BCL2
  - TCL1A
  - TWIST2
  - CDKN2A

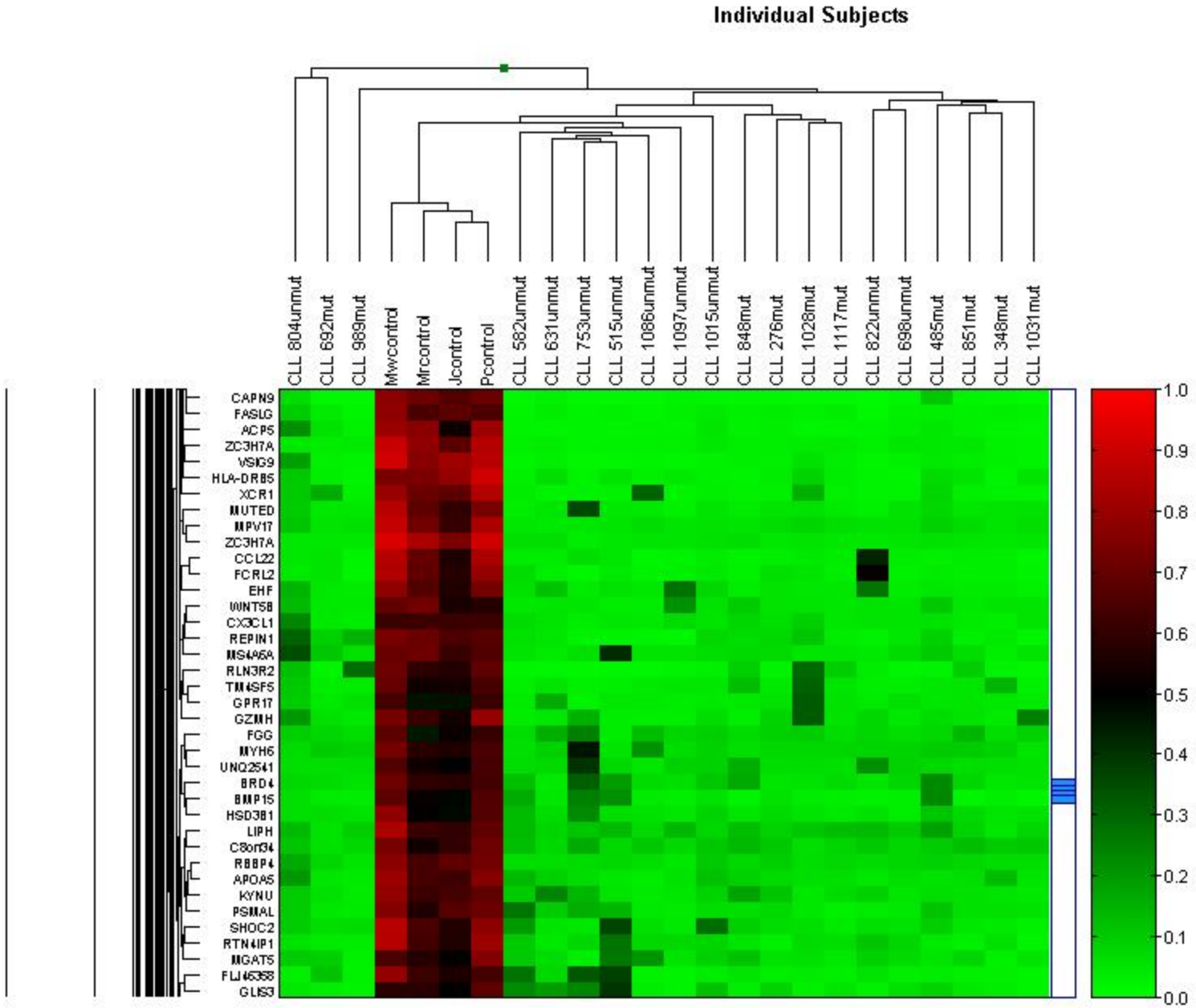
Custom Methylation Assay by Group



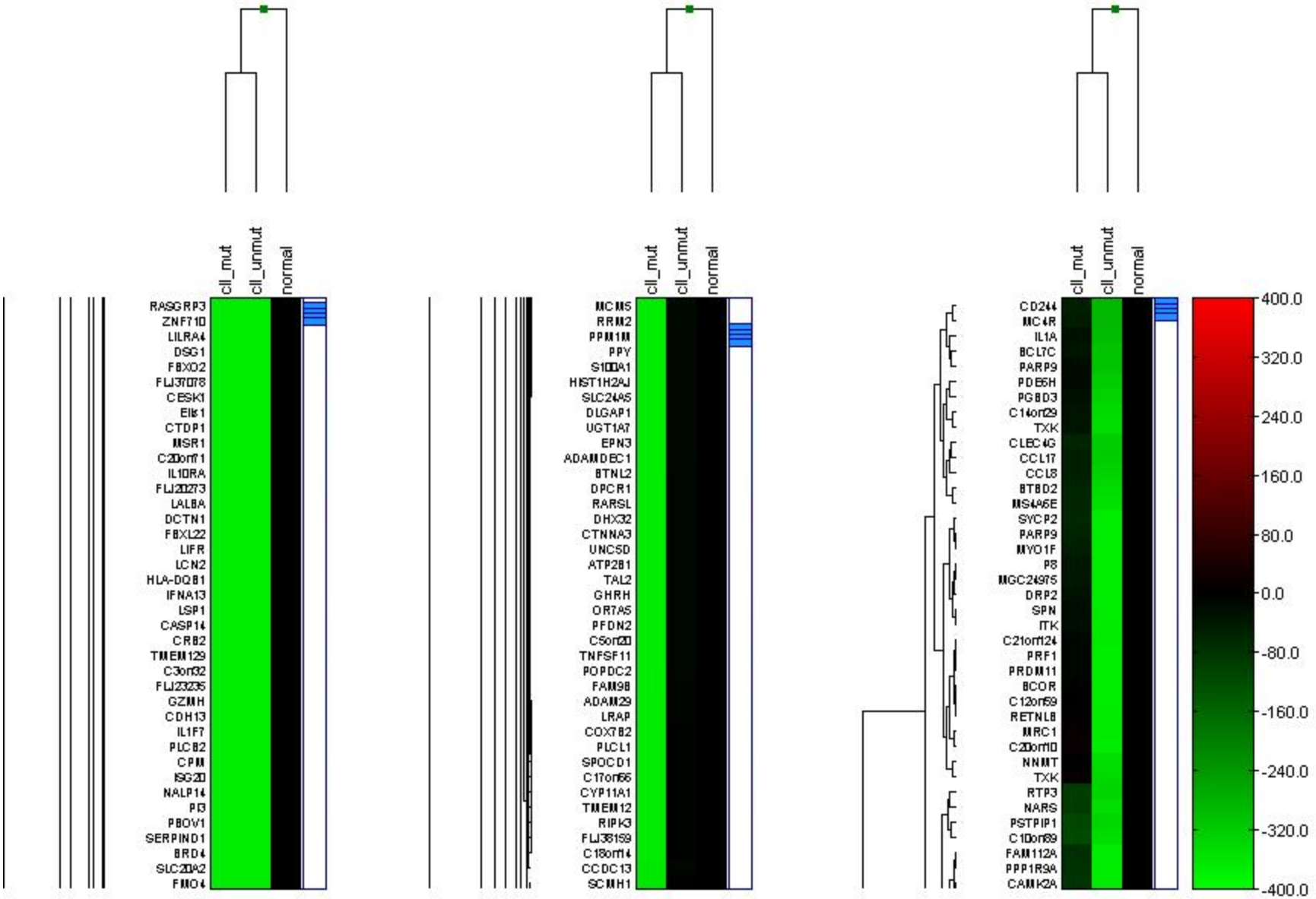
Genome-wide Methylation Assay

- HumanMethylation27 BeadChip
  - >27,000 CpG loci
  - >14,000 genes

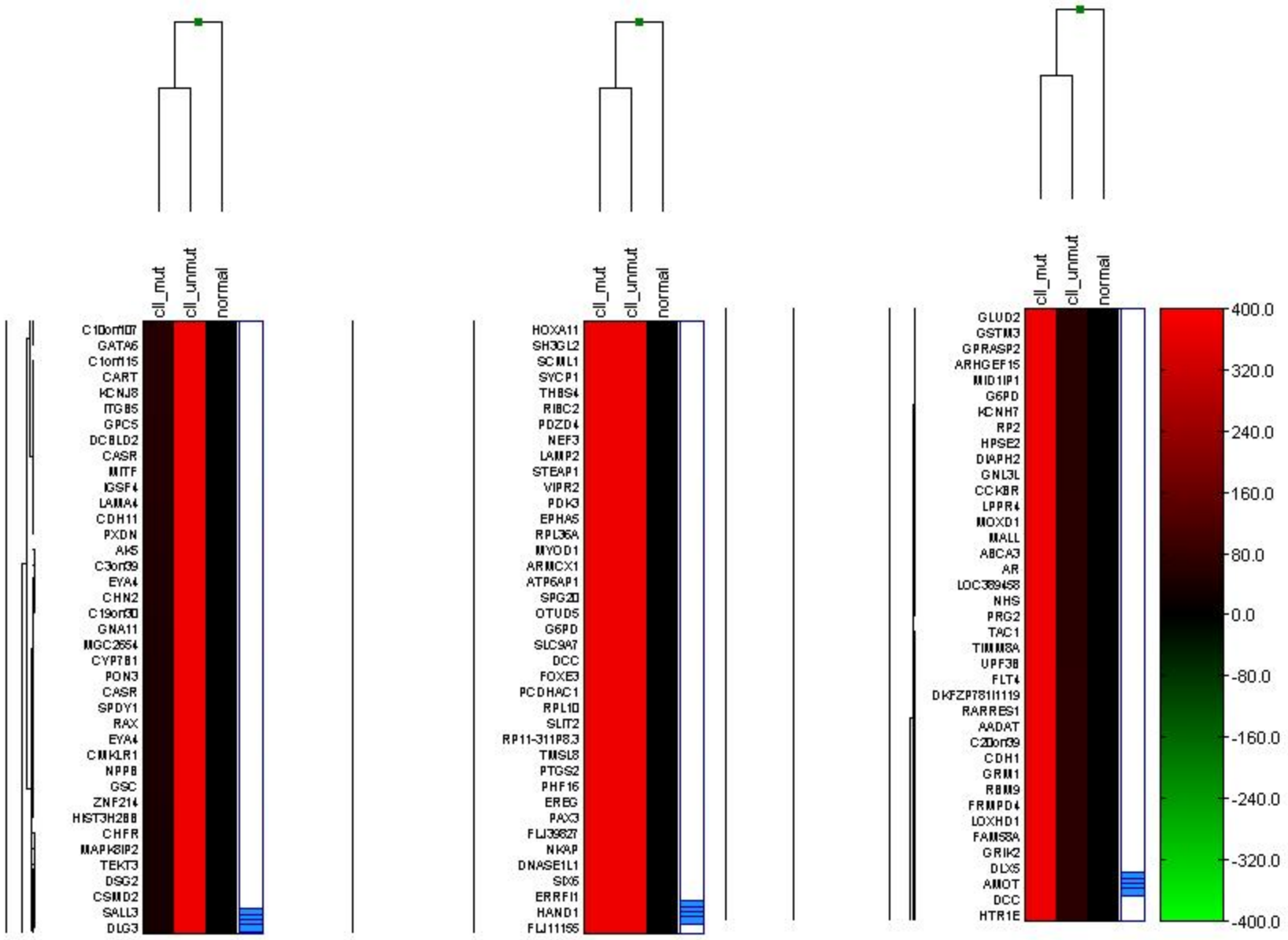



Individual Subjects
CLL 804unmut
CLL 692mut
CLL 989mut
Mwcontrol
Mrcontrol
Jcontrol
Pcontrol
CLL 582unmut
CLL 631unmut
CLL 753unmut
CLL 515unmut
CLL 1086unmut
CLL 1097unmut
CLL 1015unmut
CLL 848mut
CLL 276mut
CLL 1028mut
CLL 1117mut
CLL 822unmut
CLL 698unmut
CLL 485mut
CLL 851mut
CLL 348mut
CLL 1031mut
CAPN9
FASLG
ACP5
ZC3H7A
VSIG9
HLA-DRB5
XCR1
MUTED
MPV17
ZC3H7A
CCL22
FCRL2
EHF
WNT5B
CX3CL1
REPIN1
MS4A6A
RLN3R2
TM4SF5
GPR17
GZMH
FGG
MYH6
UNQ2541
BRD4
BMP15
HSD3B1
LIPH
C8orf34
RBBP4
APOA5
KYNU
PSMAL
SHOC2
RTN4IP1
MGAT5
FLJ45368
GLIS3
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0

Differential Scores (normalized to control)



Differential Scores (normalized to control)



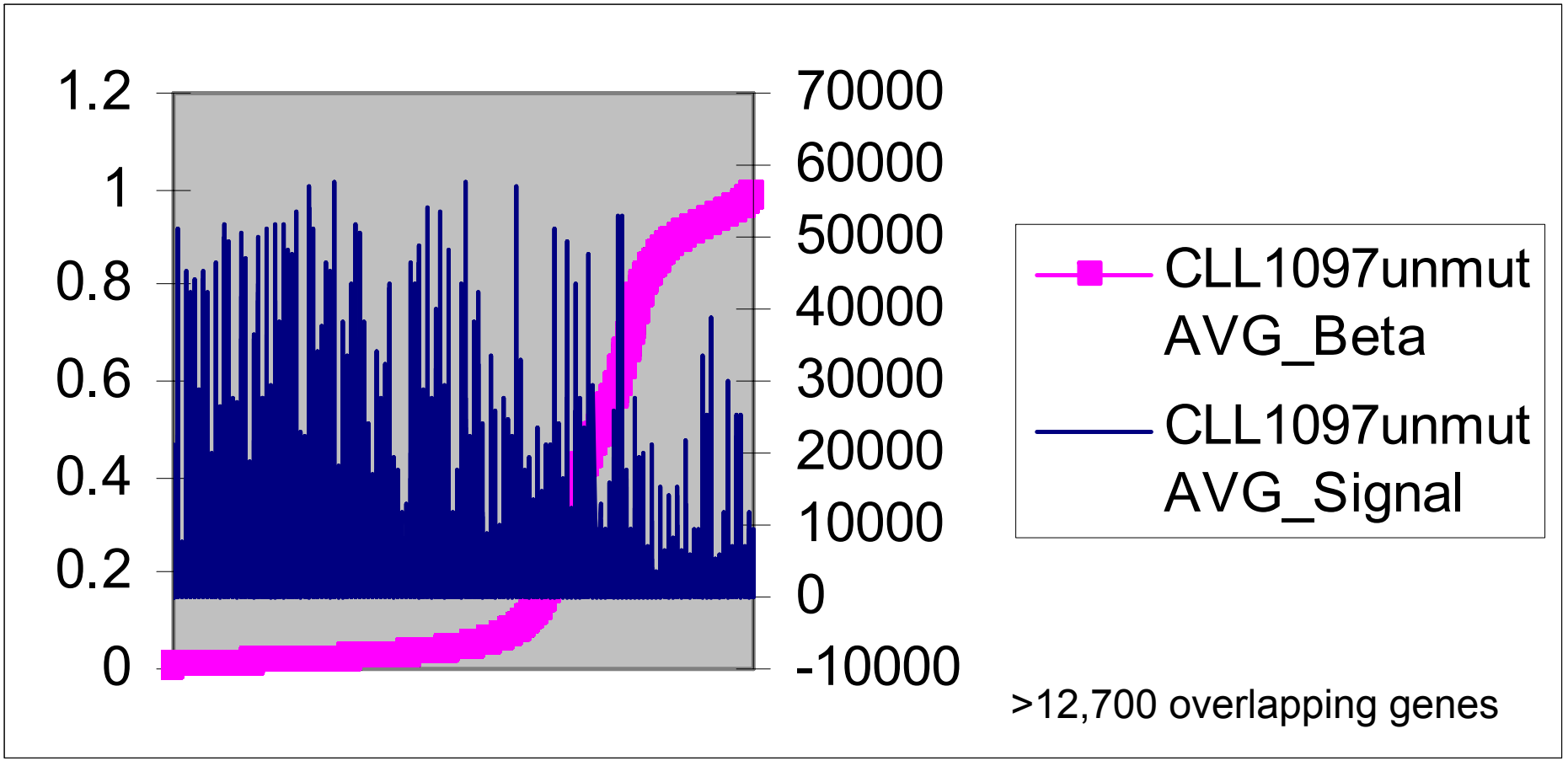
DNA Methylation

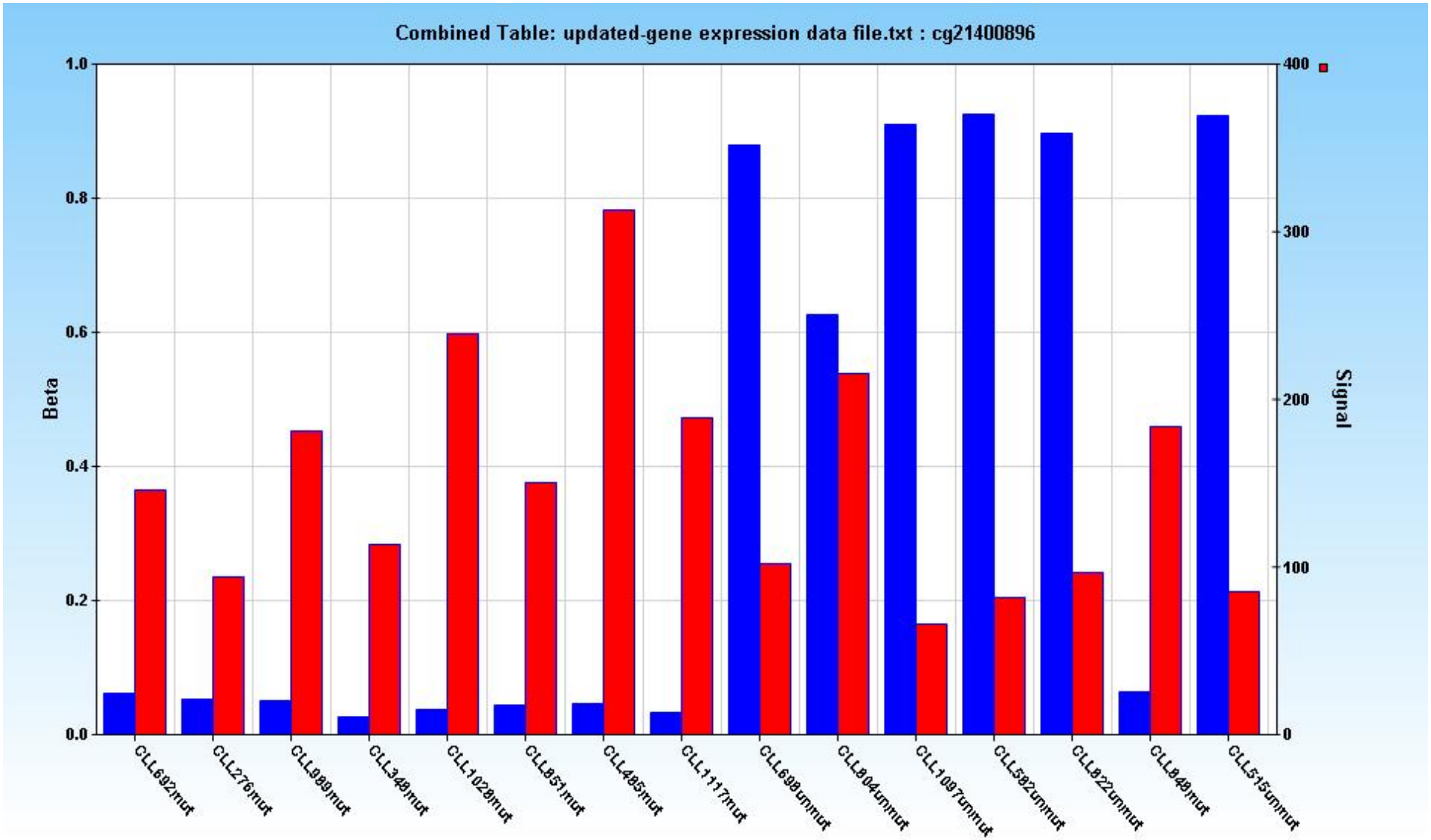
- Clear differences in DNA methylation patterns between control B cells, CLL (Vh mutated and unmutated) PBMCs
- Next step :
  - Analyze more samples in subgroups
  - Compare methylation data to gene expression results on same individuals

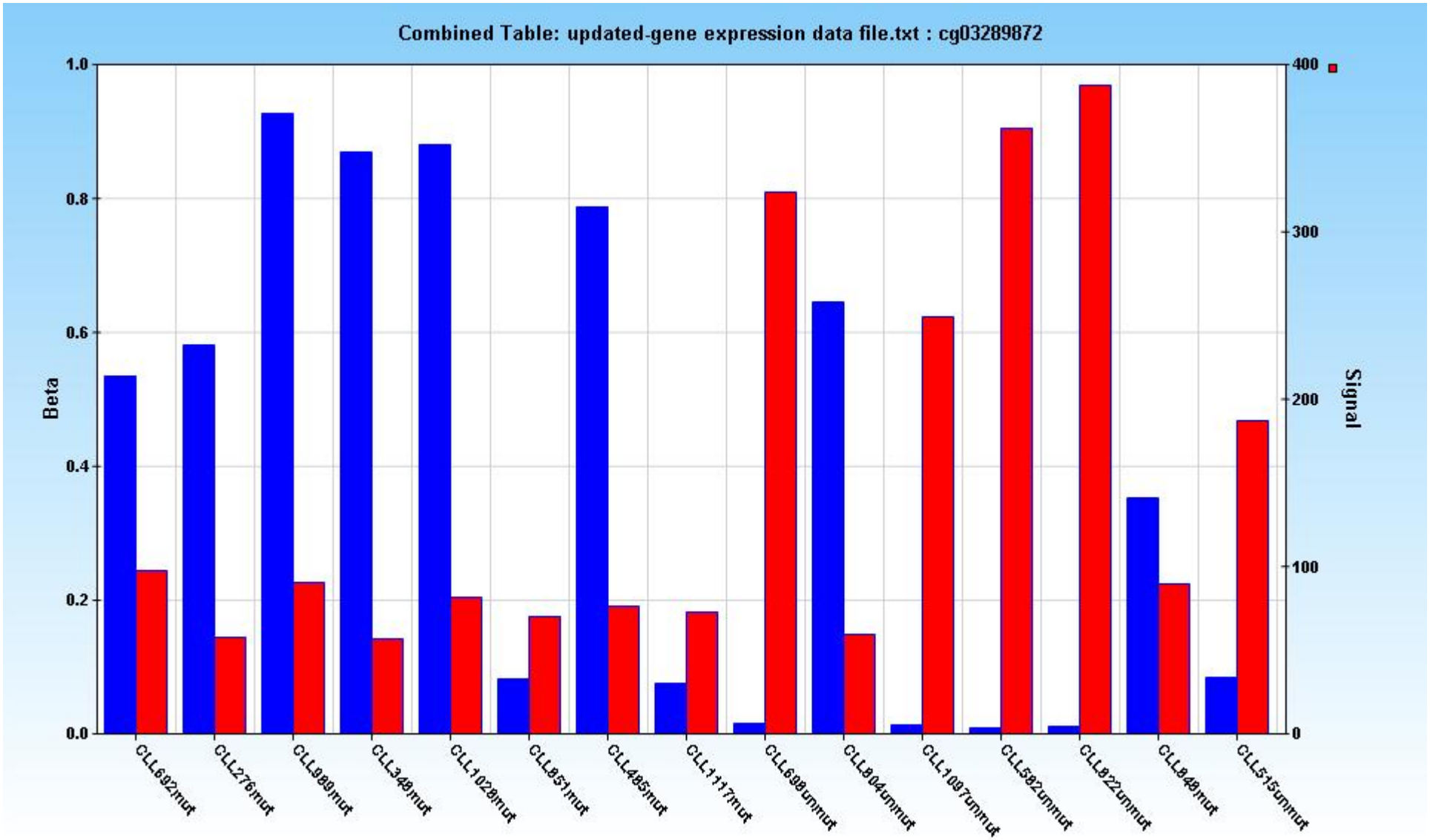
Genome wide Expression

- Human WG 6 V.3
  - > 48,000 probes
  - >25,000 genes

Methylation vs. Expression




Combined Table: updated-gene expression data file.txt : cg21400896
Beta
Signal
1.0
0.8
0.6
0.4
0.2
0.0
400
300
200
100
0
CLL692mut
CLL276mut
CLL989mut
CLL348mut
CLL1028mut
CLL851mut
CLL485mut
CLL1117mut
CLL698unmut
CLL804unmut
CLL1097unmut
CLL582unmut
CLL822unmut
CLL848mut
CLL515unmut

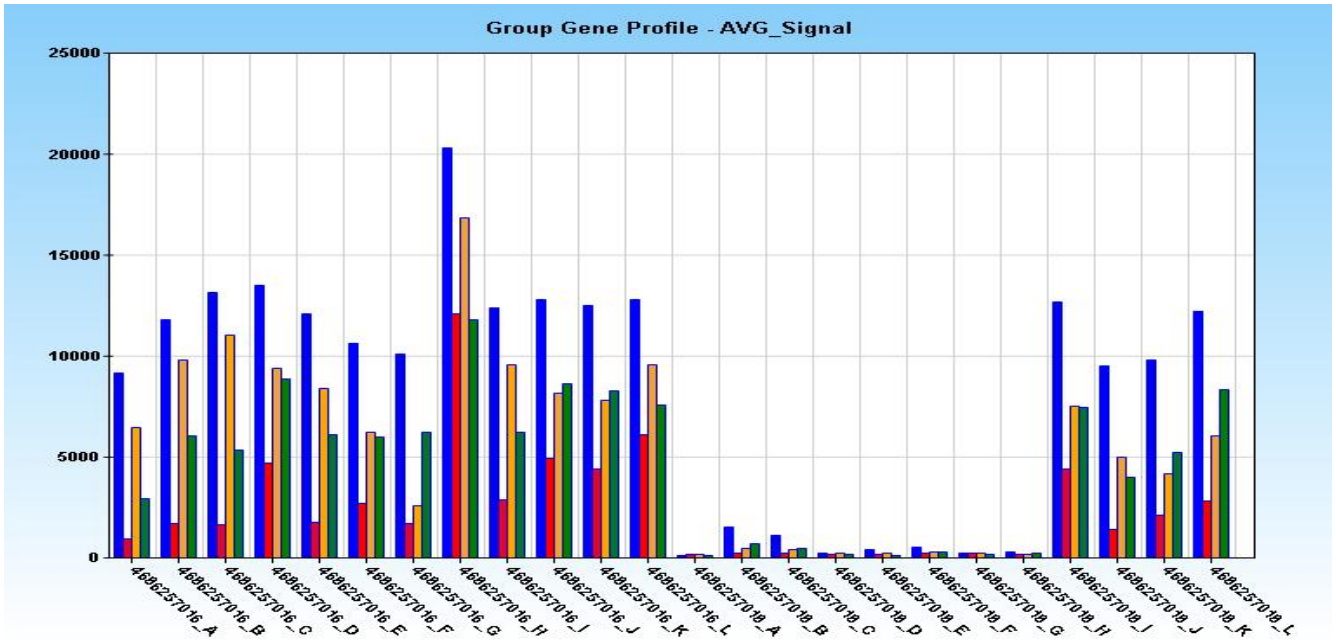

Combined Table: updated-gene expression data file.txt : cg03289872
Beta
Signal
1.0
0.8
0.6
0.4
0.2
0.0
400
300
200
100
0
CLL692mut
CLL276mut
CLL989mut
CLL348mut
CLL1028mut
CLL851mut
CLL485mut
CLL1117mut
CLL698unmut
CLL804unmut
CLL1097unmut
CLL582unmut
CLL822unmut
CLL848mut
CLL515unmut

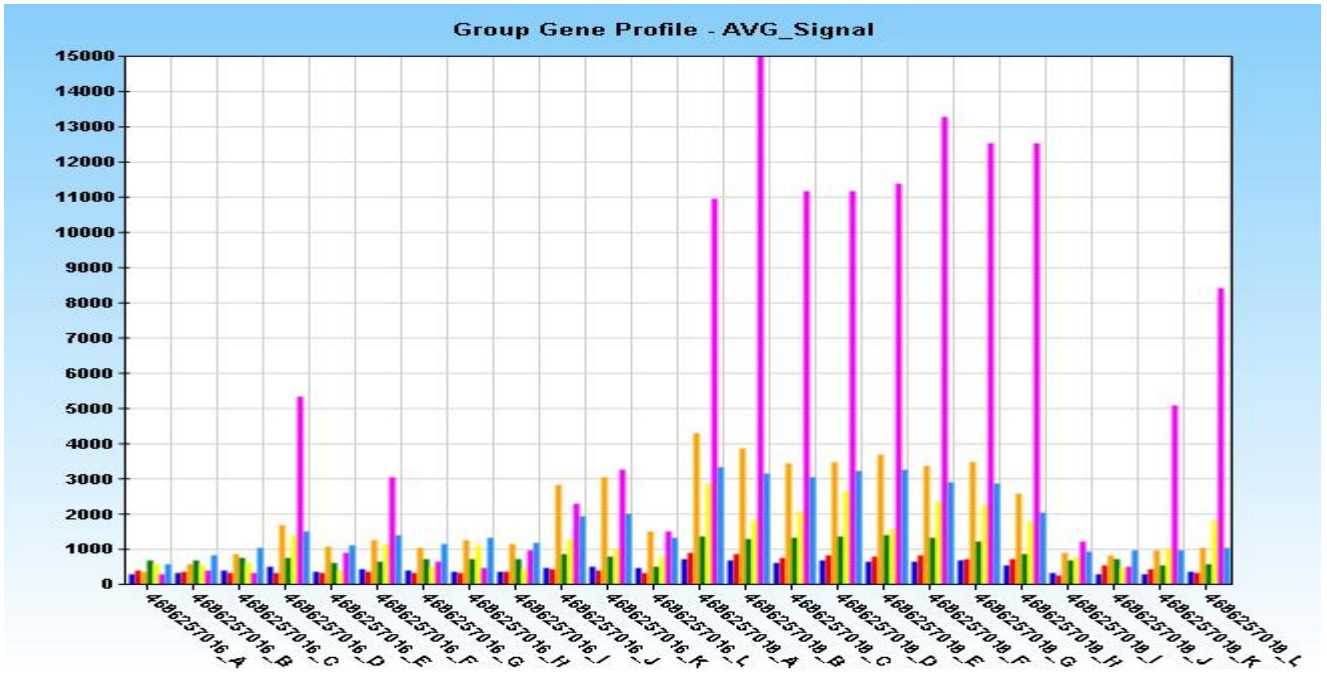
MicroRNA Expression

- Small noncoding RNAs that regulate gene expression post-transcriptionally

- Dysregulation associated with disease
  - Cancer
    - Oncogene
    - Tumor suppressor

V2 panel - 1146 microRNAs


Group Gene Profile - AVG_Signal
25000
20000
15000
10000
5000
0
4686257016_A
4686257016_B
4686257016_C
4686257016_D
4686257016_E
4686257016_F
4686257016_G
4686257016_H
4686257016_I
4686257016_J
4686257016_K
4686257016_L
4686257019_A
4686257019_B
4686257019_C
4686257019_D
4686257019_E
4686257019_F
4686257019_G
4686257019_H
4686257019_I
4686257019_J
4686257019_K
4686257019_L


Group Gene Profile - AVG_Signal
15000
14000
13000
12000
11000
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
4686257016_A
4686257016_B
4686257016_C
4686257016_D
4686257016_E
4686257016_F
4686257016_G
4686257016_H
4686257016_I
4686257016_J
4686257016_K
4686257016_L
4686257019_A
4686257019_B
4686257019_C
4686257019_D
4686257019_E
4686257019_F
4686257019_G
4686257019_H
4686257019_I
4686257019_J
4686257019_K
4686257019_L

Expanded Dataset

- CLL Clinical Trial

  (~2000 subjects, asymptomatic, within 6 mos of dx)

  - Good prognostic markers
    - no therapy, @ 3 month visits
  - Bad prognostic markers
    - no therapy, @ 3 month visits
  - Bad prognostic markers
    - therapy with close follow up

Then What?

- Methylation/Expression profiles
  - Aggressive vs benign CLL
  - Responders vs nonresponders
- Develop targeted assays
  - Expression (RNA and miRNA)
  - Methylation

Acknowledgements

- Peter Gregersen
- Anthony Liew
- Houman Khalili
- Aarti Damle
- Jan Freudenberg
- Cassie Pond
- Mike Wigler
- Diane Esposito
- Nick Chiorazzi
- Kanti Rai
- Bill Kennedy
- Joy Yan
- Erin Boyle
- Rajendra Damle
- Sophie Yancopoulos
- Naomi O'Grady
- Joe Garsetti
- Jason Harger