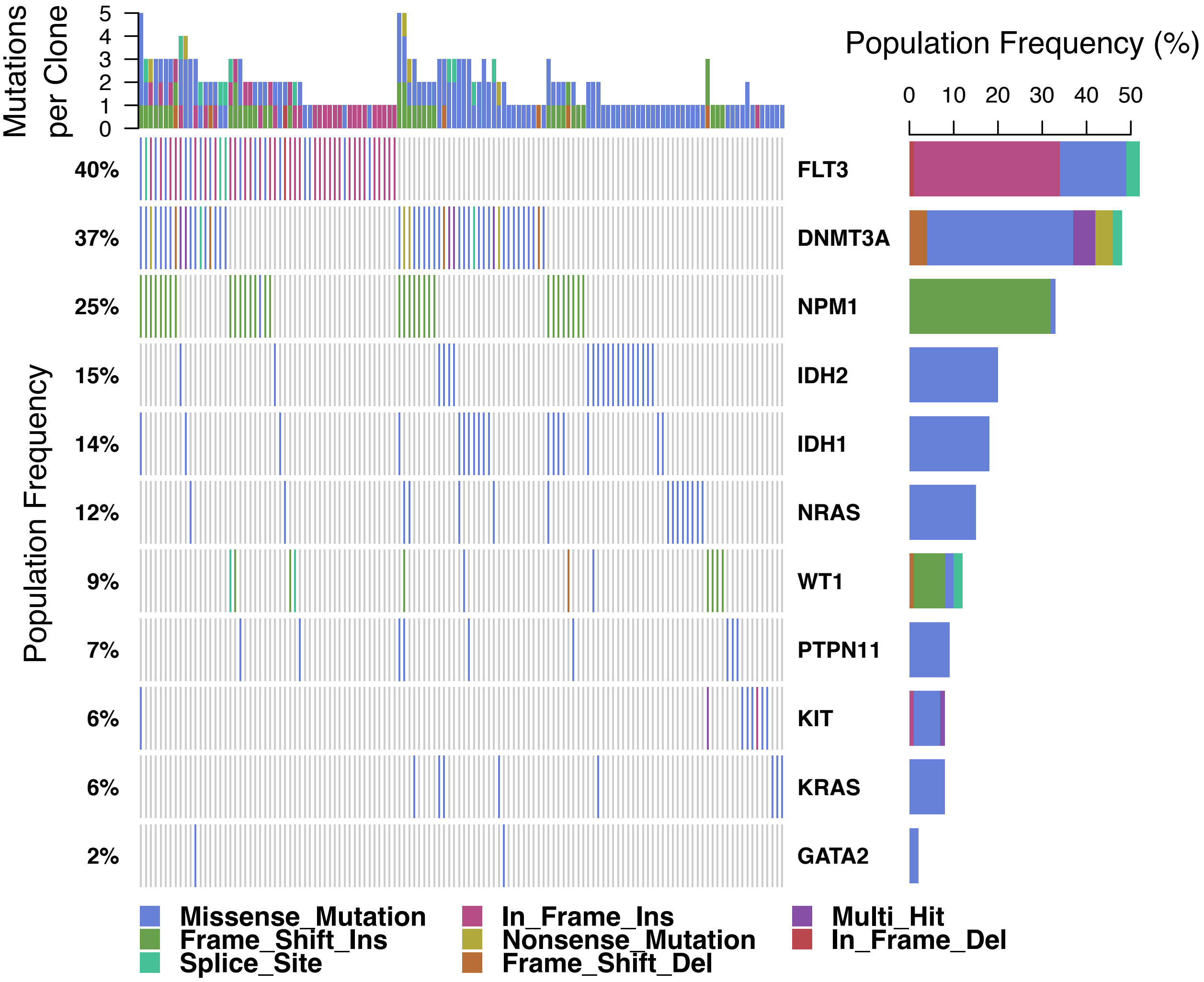
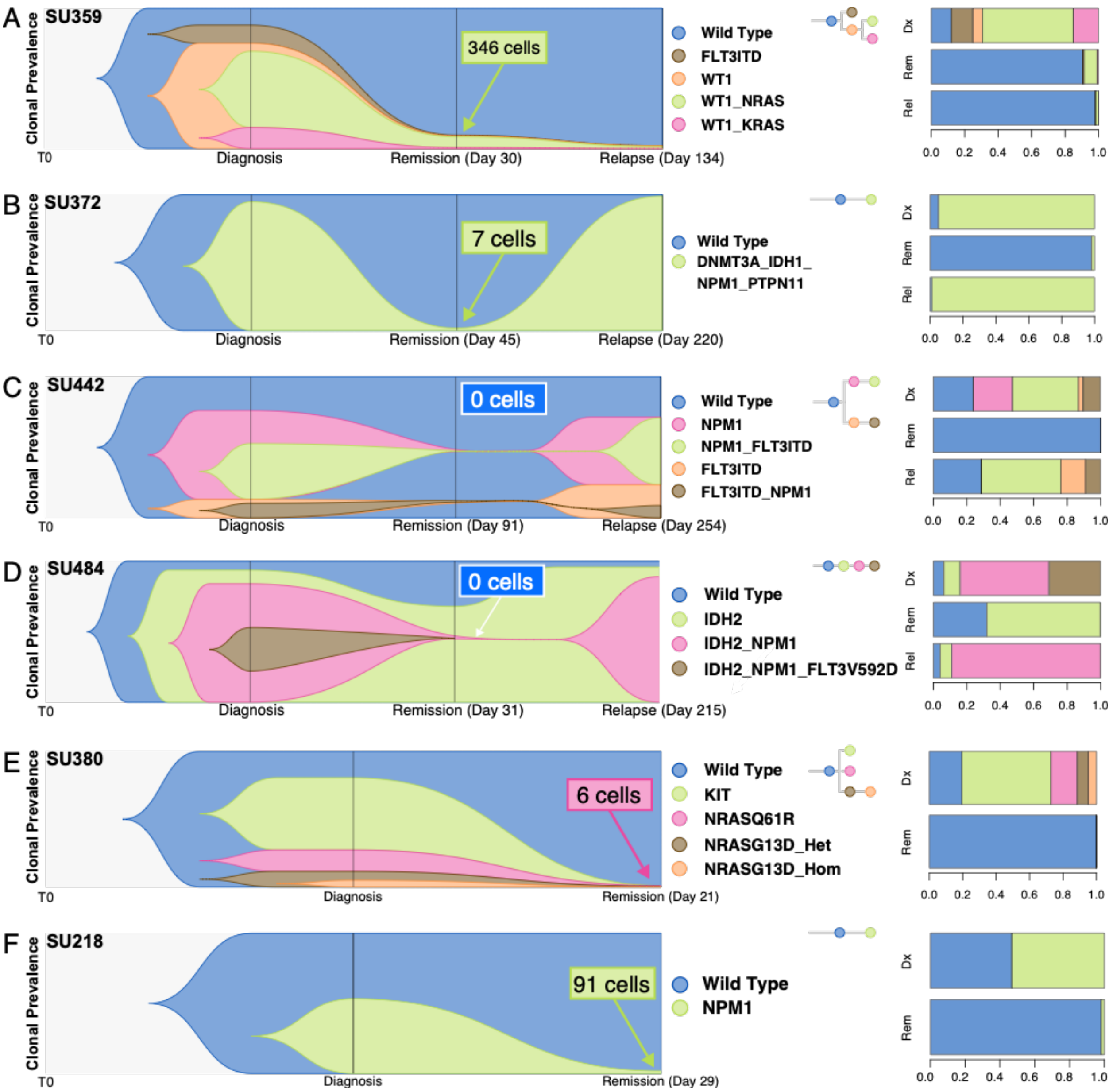
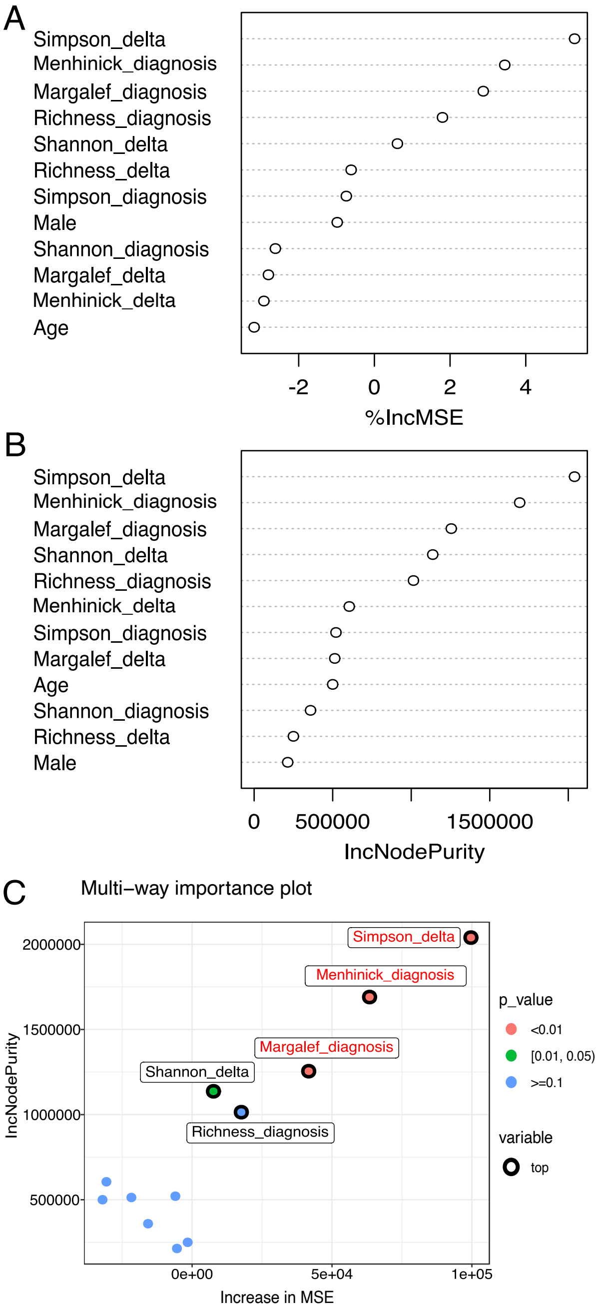
Supplementary Figures and Tables:



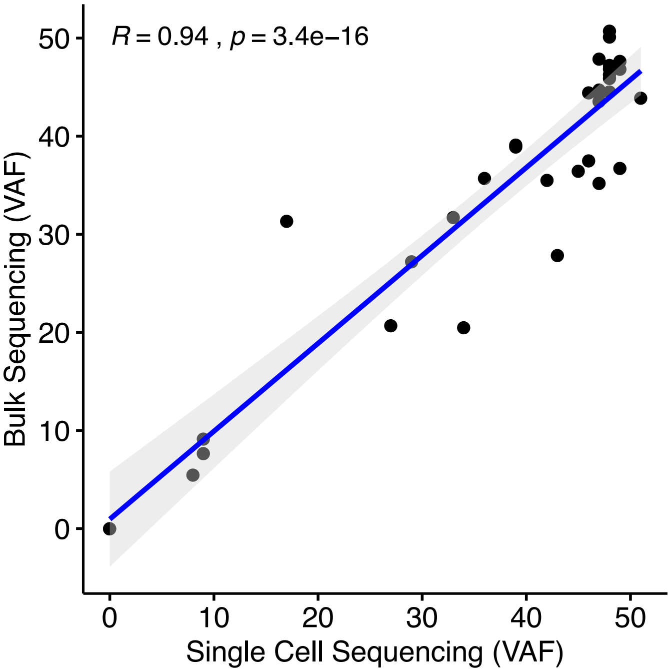
Supplemental Figure 1: Mutation landscape of AML from the TCGA. Somatic mutations are shown for 130 AML patients from the TCGA. Only cases with variants included in the AML single cell sequencing Tapestri mutation panel are presented.

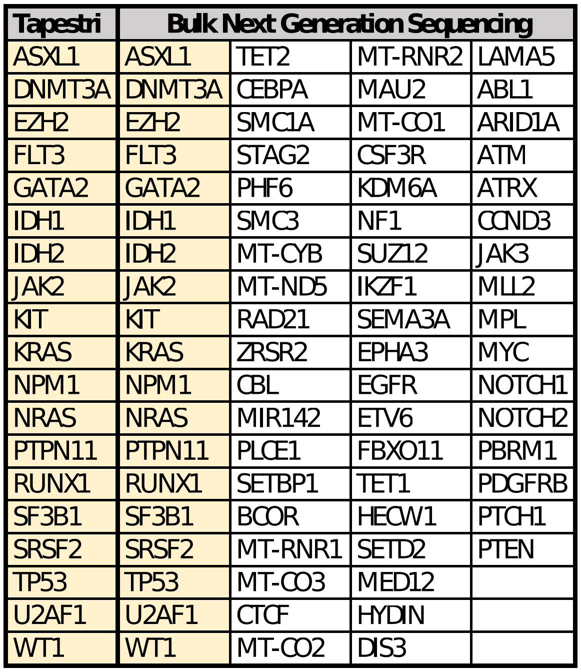


Supplemental Figure 2: Single Cell Sequencing of AML. Fishplots illustrating the diversity in clonal evolution for the remaining patients in our cohort. (A-D) Cases that relapsed. (E-F) Cases from patients that remained in remission. For each remission sample, the total number of cells identified as the dominant clone at relapse (or at diagnosis for SU218 and SU380) are indicated in the box. Dx – Diagnosis, Rem – Remission, Rel – Relapse.



Supplemental Figure 3: Evaluation of ecologic metrics in AML. Assessment of variable importance using random forest regression of RFS against age, sex, and ecologic parameters. The ecologic parameters measure either clonal richness (Menhinick, Margalef, and Richness) or diversity (Shannon and Simpson) at diagnosis (diagnosis) or the change from diagnosis to remission (delta). (A-B) Variable importance was assessed by determining the increase in mean squared error (%IncMSE) after the respective variable was permuted and the decrease in residual sum of squares by splits in the respective variable (IncNodePurity). (C) Multi-way importance plot comparing IncNodePurity and increase in MSE. Variables are colored according to significance as assessed by a one-sided binomial test. The significance represents whether the observed number of nodes where the respective variable was used for splitting exceeded the theoretical number of nodes if the variable was chosen at random. The top five variables are circled.





Tapestri	Bulk Next Generation Sequencing			
ASXL1	ASXL1	TET2	MT-RNR2	LAMA5
DNMT3A	DNMT3A	CEBPA	MAU2	ABL1
EZH2	EZH2	SMC1A	MT-CO1	ARID1A
FLT3	FLT3	STAG2	CSF3R	ATM
GATA2	GATA2	PHF6	KDM6A	ATRX
IDH1	IDH1	SMC3	NF1	CCND3
IDH2	IDH2	MT-CYB	SUZ12	JAK3
JAK2	JAK2	MT-ND5	IKZF1	MLL2
KIT	KIT	RAD21	SEMA3A	MPL
KRAS	KRAS	ZRSR2	EPHA3	MYC
NPM1	NPM1	CBL	EGFR	NOTCH1
NRAS	NRAS	MIR142	ETV6	NOTCH2
PTPN11	PTPN11	PLCE1	FBXO11	PBRM1
RUNX1	RUNX1	SETBP1	TET1	PDGFRB
SF3B1	SF3B1	BCOR	HECW1	PTCH1
SRSF2	SRSF2	MT-RNR1	SETD2	PTEN
TP53	TP53	MT-CO3	MED12	
U2AF1	U2AF1	CTCF	HYDIN	
WT1	WT1	MT-CO2	DIS3	

Supplemental Figure 4: Correlation between variants detected by single cell sequencing and bulk NGS. All variants detected by bulk NGS were recovered by SCS except for three cases where ASXL1 was missed. For the variants that were captured by both assays, the correlation between VAFs determined by NGS and SCS is plotted on the left. The list of mutations covered by the Tapestri panel and targeted bulk NGS is shown on the right. All hotspot locations covered by Tapestri were also included in the bulk sequencing panel (orange). NGS – next generation sequencing. SCS – single cell sequencing.

Supplementary Table 1: Pre-filtered single cell sequencing metrics.

Sample ID	Sample Type	Number of cells	Number of reads	Reads per Cell	Reads per Cell per Amplicon	Panel Uniformity	ADO
SU372	Diagnosis	8,092	17 M	912	18	94%	8.24%
SU372B	Remission	4,841	17 M	2,793	56	96%	10.73%
SU372C	Relapse	8,528	21 M	778	16	96%	3.34%
SU320	Diagnosis	1,855	12 M	3,845	77	96%	19.86%
SU320B	Remission	6,273	13 M	1,277	26	94%	12.55%
SU320C	Relapse	9,489	15 M	1,581	17	94%	9.95%
SU291	Diagnosis	11,243	20 M	1,292	26	96%	8.34%
SU291B	Remission	7,852	24 M	1,880	38	92%	13.32%
SU291F	Relapse	8,123	18 M	1,335	27	92%	8.26%
SU484	Diagnosis	4,429	16 M	2,714	54	96%	6.64%
SU484B	Remission	4,885	19 M	2,965	59	94%	6.56%
SU484C	Relapse	6,087	18 M	1,569	31	94%	9.23%
SU654	Diagnosis	8,335	40 M	3,929	79	94%	13.36%
SU654B	Remission	7,840	40 M	3,898	78	90%	17.33%
SU654C	Relapse	14,335	37 M	2,065	41	92%	9.56%
SU359	Diagnosis	8,727	36 M	3,200	64	92%	15.87%
SU359B	Remission 1	6,750	35 M	3,688	74	92%	12.03%
SU359C	Remission 2	5,274	34 M	4,967	99	90%	13.29%
SU067	Diagnosis	11,840	33 M	2,066	41	96%	13.01%
SU067C	Remission	4,136	29 M	6,200	124	93%	15.26%
SU067D	Relapse	3,955	31 M	6,184	124	93%	11.49%
SU674	Diagnosis	16,878	29 M	1,182	24	90%	8.68%
SU674B	Remission	12,654	31 M	1,909	38	91%	9.62%
SU674C	Relapse	11,124	28 M	1,790	36	92%	15.71%
SU442	Diagnosis	7,662	38 M	2,880	58	94%	9.94%
SU442C	Remission	6,969	30 M	3,235	65	92%	9.04%
SU442E	Relapse	8,311	34 M	2,610	52	92%	10.67%
SU353	Diagnosis	8,971	30 M	2,551	51	96%	NA
SU353B	Remission	6,442	33 M	3,571	71	95%	NA
SU353C	Relapse	11,998	36 M	2,537	51	92%	NA
SU218	Diagnosis	11,974	38 M	2,413	48	93%	NA

SU218B	Remission	6,630	32 M	3,561	71	94%	NA
SU290	Diagnosis	4,965	34 M	5,268	105	93%	16.24%
SU290B	Remission	8,726	32 M	2,704	54	94%	12.69%
SU380	Diagnosis	8,930	34 M	2,841	57	96%	12.07%
SU380B	Remission	6,983	32 M	3,317	66	92%	14.09%
SU564	Diagnosis	11,452	36 M	2,298	46	93%	NA
SU564B	Remission	7,179	36 M	3,686	74	94%	NA

ADO – allele dropout.

Supplementary Table 2:

Sample	Time	Blast %	Tissue	Multiplexed PCR results
SU067	Diagnosis	84	PB	
SU067C	Remission		BM	NPM1 PCR negative on BM
SU067D	Relapse	14	PB	
SU218	Diagnosis	35	PB	
SU218B	Remission		PB	
SU290	Diagnosis	74	PB	
SU290B	Remission		BM	
SU291	Diagnosis	35	PB	
SU291B	Remission		BM	
SU291F	Relapse	95	PB	
SU320	Diagnosis	96	BM	
SU320B	Remission		BM	
SU320C	Relapse	72	PB	
SU353	Diagnosis	66	BM	
SU353B	Remission		PB	
SU353C	Relapse	87	PB	
SU359	Diagnosis	31	PB	
SU359B	Remission 1		PB	FLT3 PCR on PB negative for ITD
SU359C	Remission 2		PB	FLT3 PCR on BM negative for ITD
SU372	Diagnosis	75	PB	
SU372B	Remission		BM	NPM1 PCR negative on BM
SU372C	Relapse	31	PB	
SU380	Diagnosis	73	PB	
SU380B	Remission		PB	
SU442	Diagnosis	27	PB	
SU442C	Remission		PB	NPM1 PCR negative on BM, FLT3 PCR negative on BM for ITD
SU442E	Relapse	45	PB	

SU484	Diagnosis	86	PB	
SU484B	Remission		PB	NPM1 PCR negative on BM
SU484C	Relapse	94	PB	
SU564	Diagnosis	49	PB	
SU564B	Remission		PB	
SU654	Diagnosis	97	PB	
SU654B	Remission		PB	
SU654C	Relapse	98	PB	
SU674	Diagnosis	91	PB	
SU674B	Remission		BM	DNMT3A PCR positive on BM, NPM1 PCR negative on BM, FLT3 PCR negative on BM for ITD
SU674C	Relapse	87	PB	

BM – bone marrow. PB – peripheral blood. ITD – internal tandem duplication.

Supplementary Table 3: Clonal cell counts at diagnosis, remission and relapse.

SU291	Diagnosis	Remission (D26)	Remission (D359)	Relapse (D535)
WT	1192	4554	4855	110
NPM1	4034	0	0	1
NPM1/IDH1/FLT3D835Y	0	0	0	2740
IDH2/NPM1	265	0	0	2
SU067	Diagnosis	Remission (D400)	Relapse (D523)	
WT	227	3176	2277	
NPM1	312	0	10	
NPM1/WT1	0	10	488	
NPM1/PTPN11	4892	0	0	
SU320	Diagnosis	Remission (D36)	Relapse (D52)	
WT	45	999	301	
IDH2	22	1568	37	
IDH2/NPM1	103	0	66	
IDH2/NPM1/KRAS	91	0	960	
IDH2/NPM1/FLT3I836del	298	0	0	
IDH2/NPM1/FLT3D835Y	34	0	0	
SU353	Diagnosis	Remission (D36)	Relapse (D62)	
WT	450	3535	165	
NPM1	41	1	0	
NPM1/FLT3ITD_Het	493	27	504	
NPM1/FLT3ITD_Hom	4050	82	7082	
SU359	Diagnosis	Remission (D30	Relapse (D134)	
WT	549	3976	4747	
FLT3ITD	594	9	0	
WT1	266	29	6	
WT1/NRAS	2498	346	86	
WT1/KRAS	693	36	0	
SU372	Diagnosis	Remission (D45)	Relapse (D22)	
WT	81	3611	14	
DNMT3A/IDH1/NPM1/PTPN11	1506	7	1318	
SU442	Diagnosis	Remission (D91)	Relapse (D254)	

WT	990	4623	856
NPM1/FLT3ITD	1636	0	1421
NPM1	972	0	1
FLT3ITD/NPM1	436	0	269
FLT3ITD	115	0	441
SU484	Diagnosis	Remission (D31)	Relapse (D215)
WT	130	1497	117
IDH2	200	3170	190
IDH2/NPM1	1092	0	2494
IDH2/NPM1/FLT3V592D	633	0	0
SU654	Diagnosis	Remission (D31)	Relapse (D248)
WT	126	1068	78
DNMT3A	70	2612	36
DNMT3A/IDH1	42	745	191
DNMT3A/IDH1/NPM1	541	1	1727
DNMT3A/IDH1 NPM1/NRAS	3361	10	2529
SU674	Diagnosis	Remission (D27)	Relapse (D416)
WT	139	4401	132
DNMT3A	132	2198	13
DNMT3A/NPM1	1313	19	86
DNMT3A/NPM1/FLT3ITD	1159	0	2442
SU218	Diagnosis	Remission (D29)	
WT	3098	4888	
NPM1	3529	91	
SU290	Diagnosis	Remission (D30)	
WT	345	6215	
GATA2_Het	511	2	
FLT3ITD8286_Het	68	0	
FLT3ITD8315_Het	37	0	
GATA2_Het/NRASG12D_Het	627	0	
GATA2_Het/FLT3ITD8286_Het	243	0	
GATA2Het/FLT3ITD8315_Het	108	0	
GATA2_Het/NRASG13D_Het	42	0	
GATA2_Het/NRASQ61R_Het	25	0	
GATA2_Het/KITD816H_Het	22	0	
GATA2_Het/NRASG12C_Het	21	0	

GATA2_Het/NRASQ61H_Het	20	0
GATA2_Het/KITN822K_Het	19	0
SU564	Diagnosis	Remission (D35)
WT	2929	7027
KIT	5934	6
SU380	Diagnosis	Remission (D21)
WT	1218	5761
KIT	3335	0
NRASQ61R	986	6
NRASG13D_Het	413	1
NRASG13D_Hom	309	0

Supplementary Table 4: Patient classifications based on relative changes in diversity at remission compared to diagnosis.

Sample	Relative Diversity at Remission
SU067	Higher
SU320	Higher
SU353	Higher
SU372	Higher
SU484	Higher
SU654	Higher
SU674	Higher
SU218	Lower
SU290	Lower
SU291	Lower
SU359	Lower
SU380	Lower
SU442	Lower
SU564	Lower

Supplemental Table 5: Data table for R-script 1: ALL.MAF

Hugo_Symbol	Chromosome	Start_Position	End_position	Variant_Classification	Variant_Type	Reference_Allele	Tumor_Seq_Allele2*	Tumor_Sample_Barcode	Protein_Change	i_transcript_name
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N	SU067dx_c2	NPM1:p.W288Cfs*12	NM_002520.6
PTPN11	12	112888198	112888198	Missense_Mutation	SNP	G	N	SU067dx_c2	PTPN11:p.A72T	NM_002834.4
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N	SU067dx_c4	NPM1:p.W288Cfs*12	NM_002520.6
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N	SU067rel_c3	NPM1:p.W288Cfs*12	NM_002520.6
WT1	11	32417946	32417947	Frame_Shift_Ins	INS	C	N	SU067rel_c3	WT1:p.R374Lfs*8	NM_024426.5

NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N	SU067rel_c4	NPM1:p.W288Cfs*12	NM_002520.6
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N	SU067rem_c3	NPM1:p.W288Cfs*12	NM_002520.6
WT1	11	32417946	32417947	Frame_Shift_Ins	INS	C	N	SU067rem_c3	WT1:p.R374Lfs*8	NM_024426.5
NPM1	5	170837546	170837547	Frame_Shift_Ins	INS	T	N	SU218dx_c2	NPM1:p.W288Cfs*12	NM_002520.6
NPM1	5	170837546	170837547	Frame_Shift_Ins	INS	T	N	SU218rem_c2	p.W288Cfs*12	NM_002520.6
GATA2	3	128202768	128202768	Missense_Mutation	SNP	C	N	SU290dx_c10		NM_032638.4
KIT	4	55599320	55599320	Missense_Mutation	SNP	G	N	SU290dx_c10	p.D816H	NM_000222.2
GATA2	3	128202768	128202768	Missense_Mutation	SNP	C	N	SU290dx_c11		NM_032638.4
NRAS	1	115258748	115258748	Missense_Mutation	SNP	C	N	SU290dx_c11	p.G12C	NM_002524.5
GATA2	3	128202768	128202768	Missense_Mutation	SNP	C	N	SU290dx_c12		NM_032638.4
NRAS	1	115256528	115256528	Missense_Mutation	SNP	T	N	SU290dx_c12	p.Q61H	NM_002524.5
GATA2	3	128202768	128202768	Missense_Mutation	SNP	C	N	SU290dx_c13		NM_032638.4
KIT	4	55599340	55599340	Missense_Mutation	SNP	T	N	SU290dx_c13	p.N822K	NM_000222.2
GATA2	3	128202768	128202768	Missense_Mutation	SNP	C	N	SU290dx_c2		NM_032638.4
NRAS	1	115258747	115258747	Missense_Mutation	SNP	C	N	SU290dx_c2	p.G12D	NM_002524.5
GATA2	3	128202768	128202768	Missense_Mutation	SNP	C	N	SU290dx_c3		NM_032638.4
FLT3	13	28608286	28608287	Frame_Shift_Ins	INS	X	N	SU290dx_c4		
GATA2	3	128202768	128202768	Missense_Mutation	SNP	C	N	SU290dx_c4		NM_032638.4
FLT3	13	28608315	28608316	Frame_Shift_Ins	INS	X	N	SU290dx_c5		
GATA2	3	128202768	128202768	Missense_Mutation	SNP	C	N	SU290dx_c5		NM_032638.4
FLT3	13	28608286	28608287	Frame_Shift_Ins	INS	X	N	SU290dx_c6		
GATA2	3	128202768	128202768	Missense_Mutation	SNP	C	N	SU290dx_c7		NM_032638.4
NRAS	1	115258744	115258744	Missense_Mutation	SNP	C	N	SU290dx_c7	p.G13D	NM_002524.5
FLT3	13	28608315	28608316	Frame_Shift_Ins	INS	X	N	SU290dx_c8		
GATA2	3	128202768	128202768	Missense_Mutation	SNP	C	N	SU290dx_c9		NM_032638.4
NRAS	1	115256529	115256529	Missense_Mutation	SNP	T	N	SU290dx_c9	p.Q61R	NM_002524.5
GATA2	3	128202768	128202768	Missense_Mutation	SNP	C	N	SU290rem_c3		NM_032638.4
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N	SU291dx_c2		NM_002520.6
IDH2	15	90631935	90631935	Missense_Mutation	SNP	G	N	SU291dx_c4		NM_002168.3
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N	SU291dx_c4		NM_002520.6
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N	SU291rel_c2		NM_002520.6
FLT3	13	28592642	28592642	Missense_Mutation	SNP	C	N	SU291rel_c3		NM_004119.2
IDH1	2	209113112	209113112	Missense_Mutation	SNP	C	N	SU291rel_c3		NM_005896.3
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N	SU291rel_c3		NM_002520.6
IDH2	15	90631935	90631935	Missense_Mutation	SNP	G	N	SU291rel_c4		NM_002168.3
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N	SU291rel_c4		NM_002520.6
IDH2	15	90631934	90631934	Missense_Mutation	SNP	C	N	SU320dx_c1	IDH2:p.R140Q	NM_002168.3
IDH2	15	90631934	90631934	Missense_Mutation	SNP	C	N	SU320dx_c3	IDH2:p.R140Q	NM_002168.3
KRAS	12	25398284	25398284	Missense_Mutation	SNP	C	N	SU320dx_c3	KRAS:p.G12V	NM_033360.3

NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N	SU320dx_c3	NPM1:p.W288Cfs*12	NM_002520.6
FLT3	13	28592634	28592635	In_Frame_Del	DEL	CATG	N	SU320dx_c4	FLT3:p.I836del	NM_004119.2
IDH2	15	90631934	90631934	Missense_Mutation	SNP	C	N	SU320dx_c4	IDH2:p.R140Q	NM_002168.3
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N	SU320dx_c4	NPM1:p.W288Cfs*12	NM_002520.6
IDH2	15	90631934	90631934	Missense_Mutation	SNP	C	N	SU320dx_c5	IDH2:p.R140Q	NM_002168.3
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N	SU320dx_c5	NPM1:p.W288Cfs*12	NM_002520.6
FLT3	13	28592642	28592642	Missense_Mutation	SNP	C	N	SU320dx_c7	FLT3:p.D835Y	NM_004119.2
IDH2	15	90631934	90631934	Missense_Mutation	SNP	C	N	SU320dx_c7	IDH2:p.R140Q	NM_002168.3
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N	SU320dx_c7	NPM1:p.W288Cfs*12	NM_002520.6
IDH2	15	90631934	90631934	Missense_Mutation	SNP	C	N	SU320rel_c1	IDH2:p.R140Q	NM_002168.3
IDH2	15	90631934	90631934	Missense_Mutation	SNP	C	N	SU320rel_c3	IDH2:p.R140Q	NM_002168.3
KRAS	12	25398284	25398284	Missense_Mutation	SNP	C	N	SU320rel_c3	KRAS:p.G12V	NM_033360.3
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N	SU320rel_c3	NPM1:p.W288Cfs*12	NM_002520.6
IDH2	15	90631934	90631934	Missense_Mutation	SNP	C	N	SU320rel_c5	IDH2:p.R140Q	NM_002168.3
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N	SU320rel_c5	NPM1:p.W288Cfs*12	NM_002520.6
IDH2	15	90631934	90631934	Missense_Mutation	SNP	C	N	SU320rem_c1	IDH2:p.R140Q	NM_002168.3
FLT3	13	28608275	28608276	Frame_Shift_Ins	INS	X	N	SU353dx_c1		
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N	SU353dx_c1		NM_002520.6
FLT3	13	28608275	28608276	Frame_Shift_Ins	INS	X	N	SU353dx_c3		
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N	SU353dx_c3		NM_002520.6
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N	SU353dx_c4		NM_002520.6
FLT3	13	28608275	28608276	Frame_Shift_Ins	INS	X	N	SU353rel_c1		
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N	SU353rel_c1		NM_002520.6
FLT3	13	28608275	28608276	Frame_Shift_Ins	INS	X	N	SU353rel_c3		
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N	SU353rel_c3		NM_002520.6
FLT3	13	28608275	28608276	Frame_Shift_Ins	INS	X	N	SU353rem_c1		
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N	SU353rem_c1		NM_002520.6
FLT3	13	28608275	28608276	Frame_Shift_Ins	INS	X	N	SU353rem_c3		
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N	SU353rem_c3		NM_002520.6
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N	SU353rem_c4		NM_002520.6
NRAS	1	115258748	115258748	Missense_Mutation	SNP	C	N	SU359dx_c2		NM_002524.5
WT1	11	32417942	32417943	In_Frame_Ins	INS	A	N	SU359dx_c2		NM_024426.5
KRAS	12	25398284	25398284	Missense_Mutation	SNP	C	N	SU359dx_c3		NM_033360.3
WT1	11	32417942	32417943	In_Frame_Ins	INS	A	N	SU359dx_c3		NM_024426.5
FLT3	13	28608286	28608287	Frame_Shift_Ins	INS	X	N	SU359dx_c4		
WT1	11	32417942	32417943	In_Frame_Ins	INS	A	N	SU359dx_c5		NM_024426.5
NRAS	1	115258748	115258748	Missense_Mutation	SNP	C	N	SU359rel_c2		NM_002524.5
WT1	11	32417942	32417943	Frame_Shift_Ins	INS	A	N	SU359rel_c2		NM_024426.5
WT1	11	32417942	32417943	Frame_Shift_Ins	INS	A	N	SU359rel_c5		NM_024426.5

NRAS	1	115258748	115258748	Missense_Mutation	SNP	C	N	SU359rem_c2		NM_002524.5
WT1	11	32417942	32417943	Frame_Shift_Ins	INS	A	N	SU359rem_c2		NM_024426.5
KRAS	12	25398284	25398284	Missense_Mutation	SNP	C	N	SU359rem_c3		NM_033360.3
WT1	11	32417942	32417943	Frame_Shift_Ins	INS	A	N	SU359rem_c3		NM_024426.5
FLT3	13	28608286	28608287	Frame_Shift_Ins	INS	X	N	SU359rem_c4		
WT1	11	32417942	32417943	Frame_Shift_Ins	INS	A	N	SU359rem_c5		NM_024426.5
DNMT3A	2	25457242	25457242	Missense_Mutation	SNP	C	N	SU372dx_c2		NM_022552.4
IDH1	2	209113112	209113112	Missense_Mutation	SNP	C	N	SU372dx_c2		NM_005896.3
NPM1	5	170837544	170837545	Frame_Shift_Ins	INS	T	N	SU372dx_c2		NM_002520.6
PTPN11	12	112926852	112926852	Missense_Mutation	SNP	C	N	SU372dx_c2		NM_002834.4
DNMT3A	2	25457242	25457242	Missense_Mutation	SNP	C	N	SU372rel_c2		NM_022552.4
IDH1	2	209113112	209113112	Missense_Mutation	SNP	C	N	SU372rel_c2		NM_005896.3
NPM1	5	170837544	170837545	Frame_Shift_Ins	INS	T	N	SU372rel_c2		NM_002520.6
PTPN11	12	112926852	112926852	Missense_Mutation	SNP	C	N	SU372rel_c2		NM_002834.4
DNMT3A	2	25457242	25457242	Missense_Mutation	SNP	C	N	SU372rem_c2		NM_022552.4
IDH1	2	209113112	209113112	Missense_Mutation	SNP	C	N	SU372rem_c2		NM_005896.3
NPM1	5	170837544	170837545	Frame_Shift_Ins	INS	T	N	SU372rem_c2		NM_002520.6
PTPN11	12	112926852	112926852	Missense_Mutation	SNP	C	N	SU372rem_c2		NM_002834.4
KIT	4	55589767	55589767	Missense_Mutation	SNP	A	N	SU380dx_c2	p.T417P	NM_000222.2
NRAS	1	115256529	115256529	Missense_Mutation	SNP	T	N	SU380dx_c3	p.Q61R	NM_002524.5
NRAS	1	115258744	115258744	Missense_Mutation	SNP	C	N	SU380dx_c4	p.G13D	NM_002524.4
NRAS	1	115258744	115258744	Missense_Mutation	SNP	C	N	SU380dx_c5	p.G13D	NM_002524.4
NRAS	1	115256529	115256529	Missense_Mutation	SNP	T	N	SU380rem_c3	p.Q61R	NM_002524.5
NRAS	1	115258744	115258744	Missense_Mutation	SNP	C	N	SU380rem_c4	p.G13D	NM_002524.4
FLT3	13	28608309	28608310	Frame_Shift_Ins	INS	X	N	SU442dx_c2		
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N	SU442dx_c2		NM_002520.6
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N	SU442dx_c3		NM_002520.6
FLT3	13	28608309	28608310	Frame_Shift_Ins	INS	X	N	SU442dx_c4		
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N	SU442dx_c4		NM_002520.6
FLT3	13	28608309	28608310	Frame_Shift_Ins	INS	X	N	SU442dx_c5		
FLT3	13	28608309	28608310	Frame_Shift_Ins	INS	X	N	SU442rel_c2		
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N	SU442rel_c2		NM_002520.6
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N	SU442rel_c3		NM_002520.6
FLT3	13	28608309	28608310	Frame_Shift_Ins	INS	X	N	SU442rel_c4		
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N	SU442rel_c4		NM_002520.6
FLT3	13	28608309	28608310	Frame_Shift_Ins	INS	X	N	SU442rel_c5		
IDH2	15	90631934	90631934	Missense_Mutation	SNP	C	N	SU484dx_c1	IDH2:p.R140Q	NM_002168.3
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N	SU484dx_c1	NPM1:p.W288Cfs*12	NM_002520.6
IDH2	15	90631934	90631934	Missense_Mutation	SNP	C	N	SU484dx_c2	IDH2:p.R140Q	NM_002168.3

FLT3	13	28608281	28608281	Missense_Mutation	SNP	A	N	SU484dx_c4	FLT3:p.V592D	NM_004119.2
IDH2	15	90631934	90631934	Missense_Mutation	SNP	C	N	SU484dx_c4	IDH2:p.R140Q	NM_002168.3
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N	SU484dx_c4	NPM1:p.W288Cfs*12	NM_002520.6
IDH2	15	90631934	90631934	Missense_Mutation	SNP	C	N	SU484rel_c1	IDH2:p.R140Q	NM_002168.3
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N	SU484rel_c1	NPM1:p.W288Cfs*12	NM_002520.6
IDH2	15	90631934	90631934	Missense_Mutation	SNP	C	N	SU484rel_c2	IDH2:p.R140Q	NM_002168.3
IDH2	15	90631934	90631934	Missense_Mutation	SNP	C	N	SU484rem_c2	p.R140Q	NM_002168.3
KIT	4	55599340	55599340	Missense_Mutation	SNP	T	N	SU564dx_c2	p.N822K	NM_000222.2
KIT	4	55599340	55599340	Missense_Mutation	SNP	T	N	SU564rel_c2	p.N822K	NM_000222.2
DNMT3A	2	25457243	25457243	Missense_Mutation	SNP	G	N	SU654dx_c2		NM_022552.4
IDH1	2	209113113	209113113	Missense_Mutation	SNP	G	N	SU654dx_c2		NM_005896.3
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N	SU654dx_c2		NM_002520.6
NRAS	1	115258744	115258744	Missense_Mutation	SNP	C	N	SU654dx_c2		NM_002524.4
DNMT3A	2	25457243	25457243	Missense_Mutation	SNP	G	N	SU654dx_c3		NM_022552.4
DNMT3A	2	25457243	25457243	Missense_Mutation	SNP	G	N	SU654dx_c4		NM_022552.4
IDH1	2	209113113	209113113	Missense_Mutation	SNP	G	N	SU654dx_c4		NM_005896.3
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N	SU654dx_c4		NM_002520.6
DNMT3A	2	25457243	25457243	Missense_Mutation	SNP	G	N	SU654dx_c5		NM_022552.4
IDH1	2	209113113	209113113	Missense_Mutation	SNP	G	N	SU654dx_c5		NM_005896.3
DNMT3A	2	25457243	25457243	Missense_Mutation	SNP	G	N	SU654rel_c2		NM_022552.4
IDH1	2	209113113	209113113	Missense_Mutation	SNP	G	N	SU654rel_c2		NM_005896.3
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N	SU654rel_c2		NM_002520.6
NRAS	1	115258744	115258744	Missense_Mutation	SNP	C	N	SU654rel_c2		NM_002524.4
DNMT3A	2	25457243	25457243	Missense_Mutation	SNP	G	N	SU654rel_c3		NM_022552.4
DNMT3A	2	25457243	25457243	Missense_Mutation	SNP	G	N	SU654rel_c4		NM_022552.4
IDH1	2	209113113	209113113	Missense_Mutation	SNP	G	N	SU654rel_c4		NM_005896.3
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N	SU654rel_c4		NM_002520.6
DNMT3A	2	25457243	25457243	Missense_Mutation	SNP	G	N	SU654rel_c5		NM_022552.4
IDH1	2	209113113	209113113	Missense_Mutation	SNP	G	N	SU654rel_c5		NM_005896.3
DNMT3A	2	25457243	25457243	Missense_Mutation	SNP	G	N	SU654rem_c2		NM_022552.4
IDH1	2	209113113	209113113	Missense_Mutation	SNP	G	N	SU654rem_c2		NM_005896.3
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N	SU654rem_c2		NM_002520.6
NRAS	1	115258744	115258744	Missense_Mutation	SNP	C	N	SU654rem_c2		NM_002524.4
DNMT3A	2	25457243	25457243	Missense_Mutation	SNP	G	N	SU654rem_c3		NM_022552.4
DNMT3A	2	25457243	25457243	Missense_Mutation	SNP	G	N	SU654rem_c4		NM_022552.4
IDH1	2	209113113	209113113	Missense_Mutation	SNP	G	N	SU654rem_c4		NM_005896.3
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N	SU654rem_c4		NM_002520.6
DNMT3A	2	25457243	25457243	Missense_Mutation	SNP	G	N	SU654rem_c5		NM_022552.4
IDH1	2	209113113	209113113	Missense_Mutation	SNP	G	N	SU654rem_c5		NM_005896.3

DNMT3A	2	25457243	25457243	Missense_Mutation	SNP	G	N	SU674dx_c2		NM_022552.4
FLT3	13	28608281	28608282	Frame_Shift_Ins	INS	X	N	SU674dx_c2		
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N	SU674dx_c2		NM_002520.6
DNMT3A	2	25457243	25457243	Missense_Mutation	SNP	G	N	SU674dx_c3		NM_022552.4
DNMT3A	2	25457243	25457243	Missense_Mutation	SNP	G	N	SU674dx_c4		NM_022552.4
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N	SU674dx_c4		NM_002520.6
DNMT3A	2	25457243	25457243	Missense_Mutation	SNP	G	N	SU674rel_c2		NM_022552.4
FLT3	13	28608281	28608282	Frame_Shift_Ins	INS	X	N	SU674rel_c2		
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N	SU674rel_c2		NM_002520.6
DNMT3A	2	25457243	25457243	Missense_Mutation	SNP	G	N	SU674rel_c3		NM_022552.4
DNMT3A	2	25457243	25457243	Missense_Mutation	SNP	G	N	SU674rel_c4		NM_022552.4
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N	SU674rel_c4		NM_002520.6
DNMT3A	2	25457243	25457243	Missense_Mutation	SNP	G	N	SU674rem_c3		NM_022552.4
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N	SU674rem_c4		NM_002520.6
* Tumor sequence was removed for patient privacy										

Supplemental Table 6: Data table for R-script 1: SAMPLE.MAF

Hugo_Symbol	Chromosome	Start_Position	End_position	Variant_Classification	Variant_Type	Reference_Allele	Tumor_Seq_Allele2*	Protein_Change	i_transcript_name	Tumor_Sample_Barcode
DNMT3A	2	25457242	25457242	Missense_Mutation	SNP	C	N		NM_022552.4	SU372dx
DNMT3A	2	25457243	25457243	Missense_Mutation	SNP	G	N		NM_022552.4	SU654dx
DNMT3A	2	25457243	25457243	Missense_Mutation	SNP	G	N		NM_022552.4	SU654dx
DNMT3A	2	25457243	25457243	Missense_Mutation	SNP	G	N		NM_022552.4	SU654dx
DNMT3A	2	25457243	25457243	Missense_Mutation	SNP	G	N		NM_022552.4	SU654dx
DNMT3A	2	25457243	25457243	Missense_Mutation	SNP	G	N		NM_022552.4	SU674dx
DNMT3A	2	25457243	25457243	Missense_Mutation	SNP	G	N		NM_022552.4	SU674dx
DNMT3A	2	25457243	25457243	Missense_Mutation	SNP	G	N		NM_022552.4	SU674dx
FLT3	13	28592634	28592635	In_Frame_Del	DEL	CATG	N	FLT3:p.I836del	NM_004119.2	SU320dx
FLT3	13	28592634	28592635	In_Frame_Del	DEL	CATG	N	FLT3:p.I836del	NM_004119.2	SU320dx
FLT3	13	28592642	28592642	Missense_Mutation	SNP	C	N	FLT3:p.D835Y	NM_004119.2	SU320dx
FLT3	13	28608281	28608281	Missense_Mutation	SNP	A	N	FLT3:p.V592D	NM_004119.2	SU484dx
FLT3	13	28608275	28608276	Frame_Shift_Ins	INS	X	N			SU353dx
FLT3	13	28608275	28608276	Frame_Shift_Ins	INS	X	N			SU353dx
FLT3	13	28608286	28608287	Frame_Shift_Ins	INS	X	N			SU359dx
FLT3	13	28608309	28608310	Frame_Shift_Ins	INS	X	N			SU442dx
FLT3	13	28608309	28608310	Frame_Shift_Ins	INS	X	N			SU442dx
FLT3	13	28608309	28608310	Frame_Shift_Ins	INS	X	N			SU442dx
FLT3	13	28608281	28608282	Frame_Shift_Ins	INS	X	N			SU674dx
FLT3	13	28608286	28608287	Frame_Shift_Ins	INS	X	N			SU290dx

FLT3	13	28608286	28608287	Frame_Shift_Ins	INS	X	N			SU290dx
FLT3	13	28608315	28608316	Frame_Shift_Ins	INS	X	N			SU290dx
FLT3	13	28608315	28608316	Frame_Shift_Ins	INS	X	N			SU290dx
GATA2	3	128202768	128202768	Missense_Mutation	SNP	C	N		NM_032638.4	SU290dx
GATA2	3	128202768	128202768	Missense_Mutation	SNP	C	N		NM_032638.4	SU290dx
GATA2	3	128202768	128202768	Missense_Mutation	SNP	C	N		NM_032638.4	SU290dx
GATA2	3	128202768	128202768	Missense_Mutation	SNP	C	N		NM_032638.4	SU290dx
GATA2	3	128202768	128202768	Missense_Mutation	SNP	C	N		NM_032638.4	SU290dx
GATA2	3	128202768	128202768	Missense_Mutation	SNP	C	N		NM_032638.4	SU290dx
GATA2	3	128202768	128202768	Missense_Mutation	SNP	C	N		NM_032638.4	SU290dx
GATA2	3	128202768	128202768	Missense_Mutation	SNP	C	N		NM_032638.4	SU290dx
GATA2	3	128202768	128202768	Missense_Mutation	SNP	C	N		NM_032638.4	SU290dx
GATA2	3	128202768	128202768	Missense_Mutation	SNP	C	N		NM_032638.4	SU290dx
IDH1	2	209113112	209113112	Missense_Mutation	SNP	C	N		NM_005896.3	SU372dx
IDH1	2	209113113	209113113	Missense_Mutation	SNP	G	N		NM_005896.3	SU654dx
IDH1	2	209113113	209113113	Missense_Mutation	SNP	G	N		NM_005896.3	SU654dx
IDH1	2	209113113	209113113	Missense_Mutation	SNP	G	N		NM_005896.3	SU654dx
IDH2	15	90631935	90631935	Missense_Mutation	SNP	G	N		NM_002168.3	SU291dx
IDH2	15	90631934	90631934	Missense_Mutation	SNP	C	N	IDH2:p.R140Q	NM_002168.3	SU320dx
IDH2	15	90631934	90631934	Missense_Mutation	SNP	C	N	IDH2:p.R140Q	NM_002168.3	SU320dx
IDH2	15	90631934	90631934	Missense_Mutation	SNP	C	N	IDH2:p.R140Q	NM_002168.3	SU320dx
IDH2	15	90631934	90631934	Missense_Mutation	SNP	C	N	IDH2:p.R140Q	NM_002168.3	SU320dx
IDH2	15	90631934	90631934	Missense_Mutation	SNP	C	N	IDH2:p.R140Q	NM_002168.3	SU320dx
IDH2	15	90631934	90631934	Missense_Mutation	SNP	C	N	IDH2:p.R140Q	NM_002168.3	SU320dx
IDH2	15	90631934	90631934	Missense_Mutation	SNP	C	N	IDH2:p.R140Q	NM_002168.3	SU484dx
IDH2	15	90631934	90631934	Missense_Mutation	SNP	C	N	IDH2:p.R140Q	NM_002168.3	SU484dx
IDH2	15	90631934	90631934	Missense_Mutation	SNP	C	N	IDH2:p.R140Q	NM_002168.3	SU484dx
KIT	4	55599320	55599320	Missense_Mutation	SNP	G	N	p.D816H	NM_000222.2	SU290dx
KIT	4	55599340	55599340	Missense_Mutation	SNP	T	N	p.N822K	NM_000222.2	SU290dx
KIT	4	55589767	55589767	Missense_Mutation	SNP	A	N	p.T417P	NM_000222.2	SU380dx
KIT	4	55599340	55599340	Missense_Mutation	SNP	T	N	p.N822K	NM_000222.2	SU564dx
KRAS	12	25398284	25398284	Missense_Mutation	SNP	C	N	KRAS:p.G12V	NM_033360.3	SU320dx
KRAS	12	25398284	25398284	Missense_Mutation	SNP	C	N	KRAS:p.G12V	NM_033360.3	SU320dx
KRAS	12	25398284	25398284	Missense_Mutation	SNP	C	N		NM_033360.3	SU359dx
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N	NPM1:p.W288Cfs*12	NM_002520.6	SU067dx
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N	NPM1:p.W288Cfs*12	NM_002520.6	SU067dx
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N	NPM1:p.W288Cfs*12	NM_002520.6	SU067dx
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N		NM_002520.6	SU291dx
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N		NM_002520.6	SU291dx

NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N	NPM1:p.W288Cfs*12	NM_002520.6	SU320dx
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N	NPM1:p.W288Cfs*12	NM_002520.6	SU320dx
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N	NPM1:p.W288Cfs*12	NM_002520.6	SU320dx
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N	NPM1:p.W288Cfs*12	NM_002520.6	SU320dx
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N	NPM1:p.W288Cfs*12	NM_002520.6	SU320dx
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N		NM_002520.6	SU353dx
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N		NM_002520.6	SU353dx
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N		NM_002520.6	SU353dx
NPM1	5	170837544	170837545	Frame_Shift_Ins	INS	T	N		NM_002520.6	SU372dx
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N		NM_002520.6	SU442dx
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N		NM_002520.6	SU442dx
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N		NM_002520.6	SU442dx
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N	NPM1:p.W288Cfs*12	NM_002520.6	SU484dx
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N	NPM1:p.W288Cfs*12	NM_002520.6	SU484dx
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N		NM_002520.6	SU654dx
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N		NM_002520.6	SU654dx
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N		NM_002520.6	SU674dx
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N		NM_002520.6	SU674dx
NPM1	5	170837546	170837547	Frame_Shift_Ins	INS	T	N	NPM1:p.W288Cfs*12	NM_002520.6	SU218dx
NRAS	1	115258748	115258748	Missense_Mutation	SNP	C	N		NM_002524.5	SU359dx
NRAS	1	115258744	115258744	Missense_Mutation	SNP	C	N		NM_002524.4	SU654dx
NRAS	1	115256528	115256528	Missense_Mutation	SNP	T	N	p.Q61H	NM_002524.5	SU290dx
NRAS	1	115256529	115256529	Missense_Mutation	SNP	T	N	p.Q61R	NM_002524.5	SU290dx
NRAS	1	115258744	115258744	Missense_Mutation	SNP	C	N	p.G13D	NM_002524.5	SU290dx
NRAS	1	115258747	115258747	Missense_Mutation	SNP	C	N	p.G12D	NM_002524.5	SU290dx
NRAS	1	115258748	115258748	Missense_Mutation	SNP	C	N	p.G12C	NM_002524.5	SU290dx
NRAS	1	115256529	115256529	Missense_Mutation	SNP	T	N	p.Q61R	NM_002524.5	SU380dx
NRAS	1	115258744	115258744	Missense_Mutation	SNP	C	N	p.G13D	NM_002524.4	SU380dx
NRAS	1	115258744	115258744	Missense_Mutation	SNP	C	N	p.G13D	NM_002524.4	SU380dx
PTPN11	12	112888198	112888198	Missense_Mutation	SNP	G	N	PTPN11:p.A72T	NM_002834.4	SU067dx
PTPN11	12	112926852	112926852	Missense_Mutation	SNP	C	N		NM_002834.4	SU372dx
WT1	11	32417942	32417943	In_Frame_Ins	INS	A	N		NM_024426.5	SU359dx
WT1	11	32417942	32417943	In_Frame_Ins	INS	A	N		NM_024426.5	SU359dx
WT1	11	32417942	32417943	In_Frame_Ins	INS	A	N		NM_024426.5	SU359dx
DNMT3A	2	25457242	25457242	Missense_Mutation	SNP	C	N		NM_022552.4	SU372rel
DNMT3A	2	25457243	25457243	Missense_Mutation	SNP	G	N		NM_022552.4	SU654rel
DNMT3A	2	25457243	25457243	Missense_Mutation	SNP	G	N		NM_022552.4	SU654rel
DNMT3A	2	25457243	25457243	Missense_Mutation	SNP	G	N		NM_022552.4	SU654rel
DNMT3A	2	25457243	25457243	Missense_Mutation	SNP	G	N		NM_022552.4	SU654rel

DNMT3A	2	25457243	25457243	Missense_Mutation	SNP	G	N		NM_022552.4	SU674rel
DNMT3A	2	25457243	25457243	Missense_Mutation	SNP	G	N		NM_022552.4	SU674rel
DNMT3A	2	25457243	25457243	Missense_Mutation	SNP	G	N		NM_022552.4	SU674rel
FLT3	13	28592642	28592642	Missense_Mutation	SNP	C	N		NM_004119.2	SU291rel
FLT3	13	28608275	28608276	Frame_Shift_Ins	INS	X	N			SU353rel
FLT3	13	28608275	28608276	Frame_Shift_Ins	INS	X	N			SU353rel
FLT3	13	28608309	28608310	Frame_Shift_Ins	INS	X	N			SU442rel
FLT3	13	28608309	28608310	Frame_Shift_Ins	INS	X	N			SU442rel
FLT3	13	28608309	28608310	Frame_Shift_Ins	INS	X	N			SU442rel
FLT3	13	28608281	28608282	Frame_Shift_Ins	INS	X	N			SU674rel
IDH1	2	209113112	209113112	Missense_Mutation	SNP	C	N		NM_005896.3	SU291rel
IDH1	2	209113112	209113112	Missense_Mutation	SNP	C	N		NM_005896.3	SU372rel
IDH1	2	209113113	209113113	Missense_Mutation	SNP	G	N		NM_005896.3	SU654rel
IDH1	2	209113113	209113113	Missense_Mutation	SNP	G	N		NM_005896.3	SU654rel
IDH1	2	209113113	209113113	Missense_Mutation	SNP	G	N		NM_005896.3	SU654rel
IDH2	15	90631935	90631935	Missense_Mutation	SNP	G	N		NM_002168.3	SU291rel
IDH2	15	90631934	90631934	Missense_Mutation	SNP	C	N	IDH2:p.R140Q	NM_002168.3	SU320rel
IDH2	15	90631934	90631934	Missense_Mutation	SNP	C	N	IDH2:p.R140Q	NM_002168.3	SU320rel
IDH2	15	90631934	90631934	Missense_Mutation	SNP	C	N	IDH2:p.R140Q	NM_002168.3	SU320rel
IDH2	15	90631934	90631934	Missense_Mutation	SNP	C	N	IDH2:p.R140Q	NM_002168.3	SU484rel
IDH2	15	90631934	90631934	Missense_Mutation	SNP	C	N	IDH2:p.R140Q	NM_002168.3	SU484rel
KRAS	12	25398284	25398284	Missense_Mutation	SNP	C	N	KRAS:p.G12V	NM_033360.3	SU320rel
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N	NPM1:p.W288Cfs*12	NM_002520.6	SU067rel
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N	NPM1:p.W288Cfs*12	NM_002520.6	SU067rel
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N		NM_002520.6	SU291rel
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N		NM_002520.6	SU291rel
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N		NM_002520.6	SU291rel
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N	NPM1:p.W288Cfs*12	NM_002520.6	SU320rel
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N	NPM1:p.W288Cfs*12	NM_002520.6	SU320rel
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N		NM_002520.6	SU353rel
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N		NM_002520.6	SU353rel
NPM1	5	170837544	170837545	Frame_Shift_Ins	INS	T	N		NM_002520.6	SU372rel
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N		NM_002520.6	SU442rel
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N		NM_002520.6	SU442rel
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N		NM_002520.6	SU442rel
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N	NPM1:p.W288Cfs*12	NM_002520.6	SU484rel
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N		NM_002520.6	SU654rel
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N		NM_002520.6	SU654rel
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N		NM_002520.6	SU674rel

NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N		NM_002520.6	SU674rel
NRAS	1	115258748	115258748	Missense_Mutation	SNP	C	N		NM_002524.5	SU359rel
NRAS	1	115258744	115258744	Missense_Mutation	SNP	C	N		NM_002524.4	SU654rel
PTPN11	12	112926852	112926852	Missense_Mutation	SNP	C	N		NM_002834.4	SU372rel
WT1	11	32417946	32417947	Frame_Shift_Ins	INS	C	N	WT1:p.R374Lfs*8	NM_024426.5	SU067rel
WT1	11	32417942	32417943	Frame_Shift_Ins	INS	A	N		NM_024426.5	SU359rel
WT1	11	32417942	32417943	Frame_Shift_Ins	INS	A	N		NM_024426.5	SU359rel
DNMT3A	2	25457242	25457242	Missense_Mutation	SNP	C	N		NM_022552.4	SU372rem
DNMT3A	2	25457243	25457243	Missense_Mutation	SNP	G	N		NM_022552.4	SU654rem
DNMT3A	2	25457243	25457243	Missense_Mutation	SNP	G	N		NM_022552.4	SU654rem
DNMT3A	2	25457243	25457243	Missense_Mutation	SNP	G	N		NM_022552.4	SU654rem
DNMT3A	2	25457243	25457243	Missense_Mutation	SNP	G	N		NM_022552.4	SU654rem
DNMT3A	2	25457243	25457243	Missense_Mutation	SNP	G	N		NM_022552.4	SU674rem
FLT3	13	28608275	28608276	Frame_Shift_Ins	INS	X	N			SU353rem
FLT3	13	28608275	28608276	Frame_Shift_Ins	INS	X	N			SU353rem
FLT3	13	28608286	28608287	Frame_Shift_Ins	INS	X	N			SU359rem
GATA2	3	128202768	128202768	Missense_Mutation	SNP	C	N		NM_032638.4	SU290rem
IDH1	2	209113112	209113112	Missense_Mutation	SNP	C	N		NM_005896.3	SU372rem
IDH1	2	209113113	209113113	Missense_Mutation	SNP	G	N		NM_005896.3	SU654rem
IDH1	2	209113113	209113113	Missense_Mutation	SNP	G	N		NM_005896.3	SU654rem
IDH1	2	209113113	209113113	Missense_Mutation	SNP	G	N		NM_005896.3	SU654rem
IDH2	15	90631934	90631934	Missense_Mutation	SNP	C	N	IDH2:p.R140Q	NM_002168.3	SU320rem
IDH2	15	90631934	90631934	Missense_Mutation	SNP	C	N	p.R140Q	NM_002168.3	SU484rem
KIT	4	55599340	55599340	Missense_Mutation	SNP	T	N	p.N822K	NM_000222.2	SU564rem
KRAS	12	25398284	25398284	Missense_Mutation	SNP	C	N		NM_033360.3	SU359rem
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N	NPM1:p.W288Cfs*12	NM_002520.6	SU067rem
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N		NM_002520.6	SU353rem
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N		NM_002520.6	SU353rem
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N		NM_002520.6	SU353rem
NPM1	5	170837544	170837545	Frame_Shift_Ins	INS	T	N		NM_002520.6	SU372rem
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N		NM_002520.6	SU654rem
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N		NM_002520.6	SU654rem
NPM1	5	170837543	170837544	Frame_Shift_Ins	INS	C	N		NM_002520.6	SU674rem
NPM1	5	170837546	170837547	Frame_Shift_Ins	INS	T	N	p.W288Cfs*12	NM_002520.6	SU218rem
NRAS	1	115258748	115258748	Missense_Mutation	SNP	C	N		NM_002524.5	SU359rem
NRAS	1	115258744	115258744	Missense_Mutation	SNP	C	N		NM_002524.4	SU654rem
NRAS	1	115256529	115256529	Missense_Mutation	SNP	T	N	p.Q61R	NM_002524.5	SU380rem
NRAS	1	115258744	115258744	Missense_Mutation	SNP	C	N	p.G13D	NM_002524.4	SU380rem
PTPN11	12	112926852	112926852	Missense_Mutation	SNP	C	N		NM_002834.4	SU372rem

WT1	11	32417946	32417947	Frame_Shift_Ins	INS	C	N	WT1:p.R374Lfs*8	NM_024426.5	SU067rem
WT1	11	32417942	32417943	Frame_Shift_Ins	INS	A	N		NM_024426.5	SU359rem
WT1	11	32417942	32417943	Frame_Shift_Ins	INS	A	N		NM_024426.5	SU359rem
WT1	11	32417942	32417943	Frame_Shift_Ins	INS	A	N		NM_024426.5	SU359rem
* Tumor sequence was removed for patient privacy										

Supplemental Table 7: Data table for R-script 3: COX.CSV

Sample ID	Age	RFS	RFS_Status	ELN2017	Male	Margalef_delta	Margalef_diagnosis	Menhinick_delta	Menhinick_diagnosis	Relapse	Richness_delta	Richness_diagnosis	Shannon_delta	Shannon_diagnosis	Simpson_delta	Simpson_diagnosis
SU067	54	514	1	1	0	-0.100906	0.34883467	-0.0011315	0.05427254	1	-1	3	-0.3686051	0.39266859	-0.1770126	0.18389383
SU218	40	2665	0	1	0	0.00381715	0.11365047	0.00377576	0.0245681	0	0	2	-0.5997762	0.69103078	-0.4619997	0.4978851
SU290	64	2552	0	2	1	-1.4553851	1.56986653	-0.259132	0.28449727	0	-11	13	-1.899891	1.90279975	-0.8034609	0.80410409
SU291	36	533	1	2	0	-0.2296809	0.34838966	-0.0243416	0.05397526	1	-2	3	-0.7039594	0.70602827	-0.4106011	0.41104011
SU320	70	177	1	2	1	-0.6846775	0.93942592	-0.2280132	0.28721348	1	-4	6	-0.7577449	1.42930665	-0.2085441	0.68438595
SU353	66	59	1	3	1	0.01846848	0.46925396	0.01234149	0.07046448	1	0	4	-0.5031805	0.65932647	-0.2759475	0.33534845
SU359	22	182	1	2	0	0.00320501	0.59283648	0.0020287	0.08845556	1	0	5	-0.9232275	1.30132652	-0.4722784	0.64829816
SU372	54	220	1	2	0	-0.0272795	0.27136193	-0.0254143	0.07528284	1	0	2	-0.1901516	0.20670906	-0.0935936	0.09800598
SU380	23	2318	0	1	1	-0.2266115	0.45755622	-0.0236889	0.06318997	0	-2	5	-1.2630526	1.27291142	-0.6444134	0.64683795
SU442	62	274	1	2	0	-0.4816806	0.60017787	-0.0637263	0.09313806	1	-4	5	-1.3848714	1.38691265	-0.7205894	0.7210218
SU484	71	224	1	3	0	-0.2876195	0.52434827	-0.066361	0.11027027	1	-2	4	-0.4743953	1.10372805	-0.1736623	0.60965202
SU564	58	1778	0	3	1	0.00287225	0.11001535	0.00260426	0.02124416	0	0	2	-0.6276331	0.63451452	-0.4408178	0.44252255
SU654	46	257	1	1	1	-0.0049356	0.60033432	-0.0031638	0.09323922	1	0	5	0.31275669	0.65909891	0.2444546	0.32285843
SU674	51	420	1	2	0	-0.1642327	0.50523091	-0.0462846	0.09545044	1	-1	4	-0.3611117	1.01667598	-0.1401364	0.58775733

Supplemental Table 8: Mutation Order Results*

Start	Stop	Clone	Subclone	Success	Trials	Binom P-Val	Q-Values	lFDR
GATA2	NRAS	GATA2	GATA2_NRAS	6	6	0.015625	0.11	0.210653296
DNMT3A	NPM1	DNMT3A	DNMT3A_NPM1	6	6	0.015625	0.11	0.210653296
IDH2_NPM1	FLT3	IDH2_NPM1	FLT3_IDH2_NPM1	4	4	0.0625	0.25	0.335704901
WT1	NRAS	WT1	NRAS_WT1	3	3	0.125	0.25	0.422157974
DNMT3A	IDH1	DNMT3A	DNMT3A_IDH1	3	3	0.125	0.25	0.422157974
DNMT3A_IDH1_NPM1	KRAS	DNMT3A_IDH1_NPM1	DNMT3A_IDH1_KRAS_NPM1	3	3	0.125	0.25	0.422157974
DNMT3A_IDH1	NPM1	DNMT3A_IDH1	DNMT3A_IDH1_NPM1	3	3	0.125	0.25	0.422157974
NPM1	WT1	NPM1	NPM1_WT1	2	2	0.25	0.29	0.565318388
GATA2	KIT	GATA2	GATA2_KIT	2	2	0.25	0.29	0.565318388
IDH2_NPM1	KRAS	IDH2_NPM1	IDH2_KRAS_NPM1	2	2	0.25	0.29	0.565318388
WT1	KRAS	WT1	KRAS_WT1	2	2	0.25	0.29	0.565318388
DNMT3A_NPM1	FLT3	DNMT3A_NPM1	DNMT3A_FLT3_NPM1	2	2	0.25	0.29	0.565318388
IDH2	NPM1	IDH2	IDH2_NPM1	4	6	0.34375	0.34	0.676327969
NPM1	FLT3	NPM1	FLT3_NPM1	4	6	0.34375	0.34	0.676327969
NPM1	PTPN11	NPM1	NPM1_PTPN11	1	1	0.5	0.41	0.877771587
NPM1	IDH1	NPM1	IDH1_NPM1	1	1	0.5	0.41	0.877771587
FLT3	GATA2	FLT3	FLT3_GATA2	2	4	0.6875	0.51	1
GATA2	FLT3	GATA2	FLT3_GATA2	2	4	0.6875	0.51	1
NPM1	IDH2	NPM1	IDH2_NPM1	2	6	0.890625	0.59	1
FLT3	NPM1	FLT3	FLT3_NPM1	2	6	0.890625	0.59	1

*Data was used to generate Figure 1B

R-Script 1 (Markdown Format): Gene Landscape

~~~~
---
title: "01gene_landscape"
output:
  pdf_document: default
---

```{r setup, include=FALSE }
knitr::opts_chunk$set(echo = TRUE)
```

# Aim: Summarize mutational landscape of SCS

Ref:
https://bioconductor.org/packages/devel/bioc/vignettes/maftools/inst/doc/maftools.html#12_pre-
compiled_tcga_maf_objects

Create MAF file based on the following:
https://docs.gdc.cancer.gov/Data/File_Formats/MAF_Format/

# Load libraries

```{r, include=FALSE}
library(maftools)
library(tidyverse)
library(TCGAmutations)
library(RColorBrewer)
```

# Set paths to relevant folders

```{r, include=FALSE}
pathtodata = '/Path_to_data/gene_landscape/data/'
pathtowd = '/Path_to_working_directory/gene_landscape/'
```

# Step 1: Create MAF file

I created data sheet using maf format from the SNPINFO files
from tapestri. I also included clinical annotation in
separate file.

The different cohorts will be based on sample timing:
1) Diagnosis 2) Remission and 3) Relapse

Load MAF files

```{r}
# setwd(pathtodata)
~~~~

~~~~
# load maf files
all.maf = paste0(pathtodata, 'all.maf')
clin = paste0(pathtodata, 'clin.txt')

# sam refers to entire sample information.
# The clone number, "_c#" was removed from
# the sample label.
# will be used for bulk reconstructed vafs

samclin = paste0(pathtodata, 'clin_sam.txt')
sam.maf = paste0(pathtodata, 'sample.maf')

all = read.maf(maf = all.maf, clinicalData = clin)
sam = read.maf(maf = sam.maf, clinicalData = samclin)
```

Upload LAML data

```{r}
#path to TCGA LAML MAF file
laml.maf = system.file('extdata', 'tcga_laml.maf.gz', package = 'maftools')
#clinical information containing survival information and histology. This is optional
laml.clin = system.file('extdata', 'tcga_laml_annot.tsv', package = 'maftools')

#Get gene list from Mbio's AML panel:
genes = read.table(file = 'mbio_amlgenes.txt')
genes = as.character(genes$V1)

# Get genes detected in our clones
genes2 = unique(dx@data$Hugo_Symbol)

lamlt = read.maf(maf = laml.maf, clinicalData = laml.clin)
lamls = subsetMaf(maf = lamlt, genes = genes, isTCGA = T, mafObj = T) # subset for genes in
MBio panel
lamls2 = subsetMaf(maf = lamlt, genes = genes2, isTCGA = T, mafObj = T) # subset for genes in
MBio panel that were detected
```

Set palettes

```{r}
# set colors
pal = brewer.pal(n = 8, name = 'Set2')
col = c("#b96d37",
        "#6780d8",
        "#b2a93d",
        "#8750a6",
        "#68a04e",
        "#b74d86",
        "#45c097",
        "#ba464e")
~~~~

~~~
names(col) = c('Frame_Shift_Del','Missense_Mutation', 'Nonsense_Mutation', 'Multi_Hit', 'Frame_Shift_Ins','In_Frame_Ins', 'Splice_Site', 'In_Frame_Del')

timecolors =  c("#9970ab","#c2a5cf","#e7d4e8")
names(timecolors)=c("Diagnosis", "Remission", "Relapse")
time_cols = list(Time = timecolors)

respcolors = c("#fff7bc", "#8da0cb")
names(respcolors)=c( "Responder", "Non-Responder" )
resp_cols = list(Response = respcolors)

deltacolors=c("#636363", "#bdbdbd")
names(deltacolors)=c("High", "Low")
delta_cols = list(Delta = deltacolors)
```

Create oncoplot (Figure 1A)

```{r}
oncoplot(maf = all, top = 11, clinicalFeatures = c('Time', 'Response', 'Delta'), colors = col, annotationColor = c(time_cols, resp_cols, delta_cols))
```
~~~

R-Script 2 (Markdown Format): Co-occurrence analysis

~~~
---
title: "Co-Occurence Analysis"
output: html_document
---

```{r setup, include=FALSE hide }
knitr::opts_chunk$set(echo = TRUE)
```

## Intro

Aim: Perform Co-occurence analysis based on DISCOVER
Ref: http://ccb.nki.nl/software/discover/doc/r/discover-intro.html
Ref: https://cran.r-project.org/web/packages/corrplot/vignettes/corrplot-intro.html
Ref: https://www.nature.com/articles/nature25795#methods

# Load libraries

```{r, include=FALSE}
library(maftools)
library(tidyverse)
library(TCGAmutations)
library(RColorBrewer)
library(discover) # mutual exclusivity test
library(corrplot) # correlation analysis
~~~

~~~
library(maftools)
library(Thermimage) # used to rotate matrices
library(pi0)
library(vioplot)
```

# Step 1: Create mut matrix

```{r}
# maf objects were created in 01gene_landscape_BA.Rmd
lamls2 =mutCountMatrix(maf = lamls2)
scs = mutCountMatrix(maf = all)
scs = scs[order(rownames(scs)),]
samscs = mutCountMatrix(maf = sam)
samscs = samscs[order(rownames(samscs)),]

write.table(scs, file = 'scs_mutation_matrix.txt', sep = '\t')
write.table(lamls2, file = 'lamls2_mutation_matrix.txt', sep = '\t')
write.table(samscs, file = 'samall_mutation_matrix.txt', sep = '\t')

```

# Step 2: Co-occurence analysis

Perform correlation analysis (Figures 1C-D)

```{r}

# Run on single cell sequencing results

s = cor(t(scs), method = 'spearman')
res1 <- cor.mtest(t(scs), conf.level = .95, method = 'spearman')
pAdj <- p.adjust(c(res1[[1]]), method = "fdr")
resAdj <- matrix(pAdj, ncol = dim(res1[[1]])[1])

corrplot(s, number.cex = 0.7, method = 'color', type = 'upper', diag = F, tl.col = 'black', p.mat =
resAdj, insig = "label_sig", sig.level = c(.01, .05, 0.1), pch.cex = .9, pch.col = "black", title = 'scs')
# adj pvalues fdr

# Run on bulk constructed vafs from scs
sm = cor(t(samscs), method = 'spearman')
res6 <- cor.mtest(t(samscs), conf.level = .95, method = 'spearman')
pAdj6 <- p.adjust(c(res6[[1]]), method = "fdr")
resAdj6 <- matrix(pAdj6, ncol = dim(res6[[1]])[1])

print('sam')
corrplot(sm, number.cex = 0.7, method = 'color', type = 'upper', diag = F, tl.col = 'black', p.mat =
resAdj6, insig = "label_sig", sig.level = c(.01, .05, 0.1), pch.cex = .9, pch.col = "black", title = 'sam')
# p-adj fdr

# Run on subsetted TCGA cohort
~~~

~~~
lamls2 = lamls2[order(rownames(lamls2)),]

l2 = cor(t(lamls2), method = 'spearman')
res5 = cor.mtest(t(lamls2), conf.level = 0.95, method = 'spearman')
pAdj5 = p.adjust(c(res5[[1]]), method = "fdr")
resAdj5 <- matrix(pAdj5, ncol = dim(res5[[1]])[1])

corrplot(l2, number.cex = 0.7, method = 'color', type = 'upper', diag = F, tl.col = 'black', p.mat =
resAdj5, insig = "label_sig", sig.level = c(.01, .05, .1), pch.cex = .9, pch.col = "black", title = 'laml2')
# FDR adjusted p-values

```
~~~

Summarizing number of co-localizing mut (Figure 1F)

~~~
```{r}

coloc = colSums(scs)
hist(coloc)
names(coloc)

# Remission samples : 218, 290, 380, 564

pattern = "218|290|380|564" # remission samples
rem_index = grepl(pattern, names(coloc))

coloc_rem = coloc[rem_index]
coloc_rel = coloc[!rem_index]

#quantify numbers

col_mut_rem = sum(coloc_rem > 1)/length(coloc_rem)
col_mut_rem

col_mut_rel = sum(coloc_rel > 1)/length(coloc_rel)
col_mut_rel

vioplot(coloc_rel,coloc_rem)
x = as.numeric(c(coloc_rel,coloc_rem))
g = as.factor(c(rep("rel", length(coloc_rel)), rep("rem", length(coloc_rem))))
kruskal.test(x,g) # 0.01254
wilcox.test(x~g) # 0.0127

library(RColorBrewer)
pal = brewer.pal(3, name = 'Paired')

fit = wilcox.test(x ~ g)
p = round(fit$p.value, 3)
~~~

~~~
vioplot(x ~ g, names = c('Relapsed','Remission'), col = pal)
lab = paste('Muations per clone')
title(lab)
legend('topright', legend = c(paste('P-val:', p)))

```
~~~

R-Script 3 (Markdown Format): Diversity analysis

~~~
---
title: "Diversity Analysis"
output: html_document
---

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
```

## Intro

Aim: Run Clonal (ie species) Richness and Diversity analysis using Vegan
Ref: http://www.flutterbys.com.au/stats/tut/tut13.2.html
Ref: Nature Reviews Cancer volume 17, pages 605–619 (2017)

# Load libraries

```{r, include=FALSE}
library(RColorBrewer)
library(pi0)
library(vioplot)
library(vegan)
library(survival)
library(survminer)
library(tidyverse)
library(coxphf)
library(plyr)
```

###########################
## DATA PROCESSING
###########################

Upload and reformat data

Set Paths to Folders

```{r}
datapath = '/Path_to_data/phylo_diversity_analysis/data'
wdpath = '/Path_to_working_directory/phylo_diversity_analysis'
```
~~~

~~~
# Diversity Analysis

```{r}
#Upload
setwd(datapath)
# cellcounts.csv contains cell counts for each clone in each sample using suplemental table 3.
cellcounts = read.csv(file = 'cellcounts.csv', header = T)
rownames(cellcounts) <- cellcounts$SampleID

```



############################
# Step 1: calculate richness
############################

```{r}

data = cellcounts[,-2] # remove column relapse designation

rich=ddply(data, ~SampleID, function(x) {
  data.frame(richness=sum(x[-1]>0))
})

# step 1a: normalize richness score for number of individuals (as richness can increase if you sample more individuals)

n = apply(data[,-1]>0,1,sum)
N = apply(data[,-1],1,sum)

# Menhinick's index
men = function(x) {
  sum(x>0)/sqrt(sum(x))
}

rich_men =  ddply(cellcounts, ~SampleID, function(x){
  data.frame(Menhinick=men(x[-1]))
})

# Margalef's Index

marg = function(x) {
  (sum(x>0)-1)/log(sum(x))
}

rich_marg = ddply(cellcounts, ~SampleID, function(x){
  data.frame(Margalef = marg(x[-1]))
})

# Combine All Richness Indices
~~~

~~~
richness = cbind(rich,Menhinick=rich_men[,2],Margalef=rich_marg[,2])
richness

plot(richness$richness,richness$Menhinick)
fit=lm(richness$richness~richness$Menhinick)
summary(fit)

plot(richness$richness,richness$Margalef)
fit=lm(richness$richness~richness$Margalef)
summary(fit)

```


Richness and Menchinick index are less correlated, so will use those two

##############################
# Step 2: Calculate Diversity
##############################

Diversity takes into account species richness and eveness

```{r}

shan = ddply(cellcounts, ~SampleID, function(x) {
  data.frame(Shannon = diversity(x[-1], index = 'shannon'))
})

simp = ddply(cellcounts, ~SampleID, function(x) {
  data.frame(Simpson = diversity(x[-1], index = 'simpson'))
})

diversity = cbind(shan,Simpson = simp$Simpson)

biodiversity = cbind(richness, diversity[,2:3], Relapse = cellcounts$Relapse)
rownames(biodiversity) = biodiversity[,1]

# write.csv(biodiversity, file = 'biodiversity.csv')


```

Subset Biodiversity data into Dx and Delta values

```{r, include=FALSE}
setwd(datapath)

# Get dx samples
# identify cells with '.' which indicates non dx sample.
# for example; a dx sample would be SU654 and a rem or
# relapse sample would be labeled SU654.B
~~~

~~~
rel_index = grepl(pattern = '[.]',rownames(biodiversity))
biod_dx = biodiversity[-rel_index==FALSE,]
biod_dx = biod_dx[,-1]

# Deltas were calculated in excel using the above export:
# 'biodiversity_dx.csv'
# Now we will subset for the dx samples which contains
# delta: remission - diagnosis

biod_delta = read.csv(file = 'biodiversity_delta.csv', header = T, row.names = 1)

```


#Upload data for cox multivariate regression

```{r, include=FALSE}
setwd(datapath)
# cox file has both biodiversity metrics (dx and delta)
surv = read.csv(file = 'cox.csv', header = T, row.names = 1)

```

## Create Plots

```{r}

pal = brewer.pal(3, name = 'Paired')

setwd(wdpath)

# Comparing Richness/Diversity at Dx

#Plot all Dx parameters
names = names(biod_dx[1:6])

for (i in 1:length(names)){
  lab1 = paste(names[i])
  fit = wilcox.test(biod_dx[,i] ~ Relapse, data = biod_dx)
  p = round(fit$p.value, 3)
  #svg(file = paste0(lab1,'_dx.svg'))#, width = 569, height = 646)
  boxplot(biod_dx[,i] ~ Relapse, data = biod_dx, names = c('Remission','Relapsed'), col = pal)
  title(paste(lab1,'Index at Diagnosis'))
  legend('topleft', legend = c(paste('P-val:', p)))
  #dev.off()
}

```
~~~

~~~
# Comparing Change in Richness/Diversity

```{r}
names = names(biod_delta[1:6])

for (i in 1:length(names[])){
  lab1 = paste(names[i])
  fit = wilcox.test(biod_delta[,i] ~ Relapse, data = biod_delta)
  p = round(fit$p.value, 3)
  #svg(filename = paste0(lab1,'_delta.svg'), width = 569, height = 646)
  boxplot(biod_delta[,i] ~ Relapse, data = biod_delta, names = c('Remission','Relapsed'), col =
pal)
  lab = paste('Change in', lab1, 'Index at Remission')
  title(lab)
  legend('topleft', legend = c(paste('P-val:', p)))
  #dev.off()
}
```




##############################
# Step 3: Coxph multi-variate
##############################

```{r}
# Fit survival data using KM method; survival here will be RFS
surv_object = Surv(time = surv$RFS, event = surv$RFS_Status)

# Create categorial variables:
surv <- surv %>% mutate('Menhinick_Level' = ifelse(Menhinick_diagnosis >
median(surv$Menhinick_diagnosis), "High", "Low"))
surv$Menhinick_Level <- factor(surv$Menhinick_Level)

surv <- surv %>% mutate('Simpson_Delta_Level' = ifelse(Simpson_delta >
median(surv$Simpson_delta), "High", "Low"))
surv$Simpson_Delta_Level <- factor(surv$Simpson_Delta_Level)

surv = surv %>% mutate(Simpson_delta_group = ifelse(Simpson_delta < -0.425709, "Low",
"High"))
surv$Simpson_delta_group = factor(surv$Simpson_delta_group)

surv = surv %>% mutate(Simpson_delta_group2 = ifelse(Simpson_delta <
median(surv$Simpson_delta), "Low", "High"))
surv$Simpson_delta_group2 = factor(surv$Simpson_delta_group2)

surv$ELN2017 <- factor(surv$ELN2017)
surv$Male <- factor(surv$Male)
~~~

~~~
# Event Free Survival
# Fit survival curves
fit.RFS1 <- survfit(surv_object ~ 1, data = surv)

# Draw Relapse free survival for all patients
#svg(filename = 'surv_all.svg')#, width = 561, height = 531)
ggsurvplot(fit.RFS1, palette = "jco", conf.int = FALSE, main = "Event Free Survival (m)", surv.median.line = "h")
#dev.off()
```


# Fit COXP model
```{r}
# Fit a Cox proportional hazards model

fit.coxph3 <- coxph(surv_object ~ Menhinick_Level + Simpson_delta_group2 + Age + Male + ELN2017, data = surv)
ggforest(fit.coxph3, data = surv)

```

####################################
# Step 4: RF Analysis for Reponse
####################################

```{r}


# Remove RFS_Status and Relapse category from data frame

surv_RFS2 = surv[,!c(colnames(surv) == 'RFS_Status')]
surv_RFS2 <- surv_RFS2[,!c(colnames(surv_RFS2) == 'Relapse')][1:14] # also removing categorical values

library(randomForest)
bag.aml = randomForest(RFS ~ ., surv_RFS2, type = 'regression', importance = T)
summary(bag.aml)
importance(bag.aml)
varImpPlot(bag.aml)

#Understand RF model
#ref:
https://cran.rstudio.com/web/packages/randomForestExplainer/vignettes/randomForestExplainer.html

library(randomForestExplainer)
min_depth_frame = min_depth_distribution(bag.aml)

plot_min_depth_distribution(min_depth_frame, mean_sample = 'all_trees', k = 15)
~~~

~~~
importance_frame = measure_importance(bag.aml)

# pdf('bag.features3.pdf', width = 6, height =4)
plot_multi_way_importance(importance_frame, x_measure = "mse_increase", y_measure =
"node_purity_increase", size_measure = "p_value", no_of_labels = 5)
# dev.off()

```

####################################
# Step 5: RF Analysis for Response
####################################

```{r}
surv2 = surv %>% mutate(Simpson_delta_group2 = ifelse(Simpson_delta <
median(surv$Simpson_delta), "Low", "High"))
surv2$Simpson_delta_group = factor(surv2$Simpson_delta_group)
surv2$Simpson_delta_group2 = factor(surv2$Simpson_delta_group2)

# Draw RFS
surv_object = Surv(time = surv2$RFS, event = surv2$RFS_Status)
fit.RFS2 <- survfit(surv_object ~ Simpson_delta_group, data = surv2)
fit.RFS2 <- survfit(surv_object ~ Simpson_delta_group2, data = surv2)
#svg(filename = 'surv_simpcutoff.svg')#, width = 561, height = 531)
ggsurvplot(fit.RFS2, palette = "jco", conf.int = FALSE, title = "Relapse Free Survival",
surv.median.line = "h", xlab = 'Time (d)', ylab = 'RFS', pval = T)
#dev.off()
```
~~~