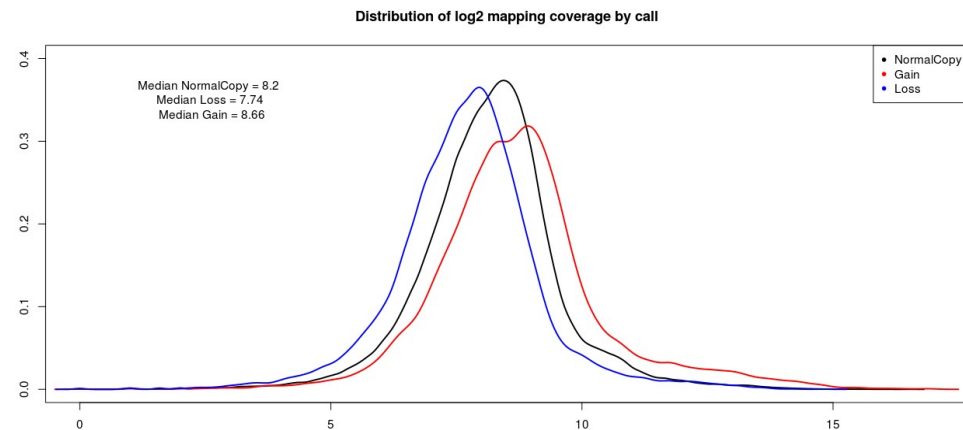


RobustCNV

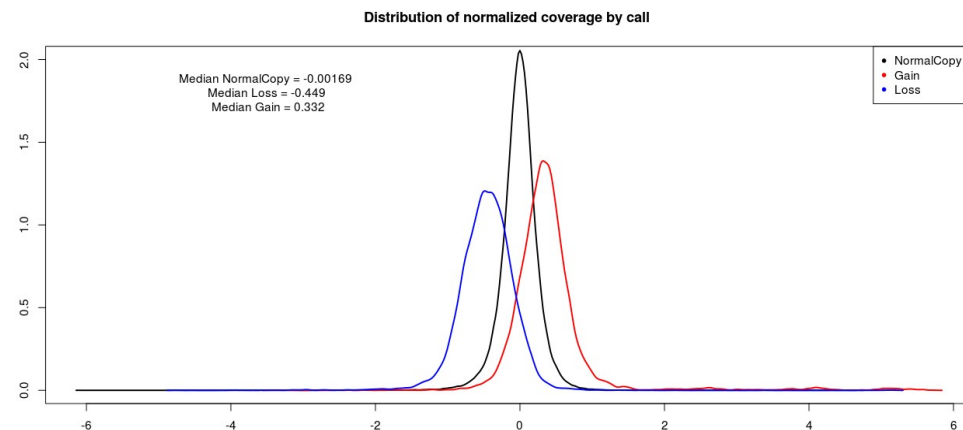
Description for Collaborators

CNV detection overview

- Little signal in raw data
 - Normalization removes systematic noise
- Segmentation breaks up intervals into groups
- Gain & loss calls made with sample-specific upper and lower threshold which is calculated based on post-normalization variability.



Normalization



RobustCNV

- Normalization
 - Robust regression against PON
 - Iteratively re-weighted least squares → robust to outliers (CNVs) as compared to ordinary least squares
 - Explicit GC normalization step removes most remaining GC bias when PON is poor
- Segmentation
 - DNACopy – Circular Binary Segmentation (Olshen and Venkatraman, 2004)
- Calling
 - Adaptive cutoff based on noise in sample.

Normalization: Stage 1

Robust Regression against PON

	Tumor	weight	PON1	PON2	PON3	...	PON _N
interval1	843	w1	547	603	579	...	14801
interval2	14995	w2	15670	12014	10766	...	11831
interval3	10930	w3	16833	7806	7985	...	5233
interval4	5645	w4	8460	4167	4699	...	2543
interval5	1510	w5	1986	1365	1239	...	13067
interval6	31967	w6	42245	22740	23146	...	8287
interval7	8643	w7	14415	8444	11162	...	7623
interval8	12025	w8	14776	11032	10553	...	10771
interval9	15936	w9	21677	14401	13054	...	20250
interval10	9068	w10	13100	7871	7234	...	9401
interval11	37790	w11	46771	36182	28392	...	8565
interval12	20440	w12	23731	17198	14487	...	9881
interval13	3627	w13	4669	1910	2042	...	19023
interval14	22654	w14	29845	18415	19862	...	21193
...	22124
interval _N	17684	w _N	24969	18085	13307	...	32116

CNV sample
(observed)

Estimate w_i with RLM

minimize "e" with RLM

$$y = \beta_1 * PON_1 + \beta_2 * PON_2 + \dots \beta_n * PON_n + e$$

$$\hat{y} = \hat{\beta}_1 * PON_1 + \hat{\beta}_2 * PON_2 + \dots \hat{\beta}_n * PON_n$$

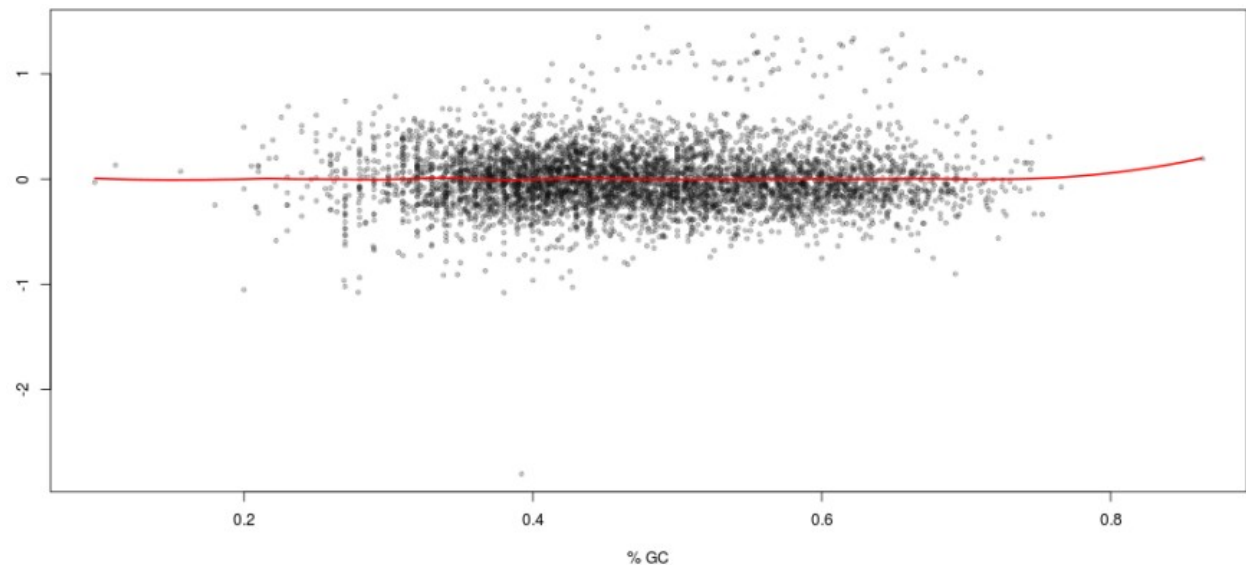
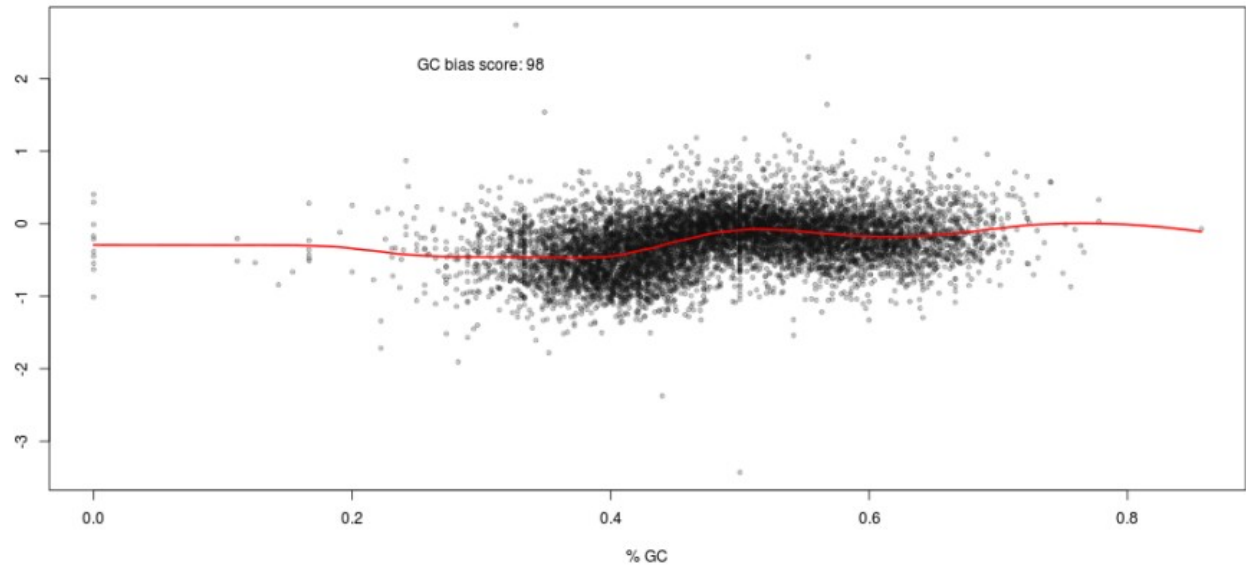
predicted

$$\text{Normalized} = y - \hat{y}$$

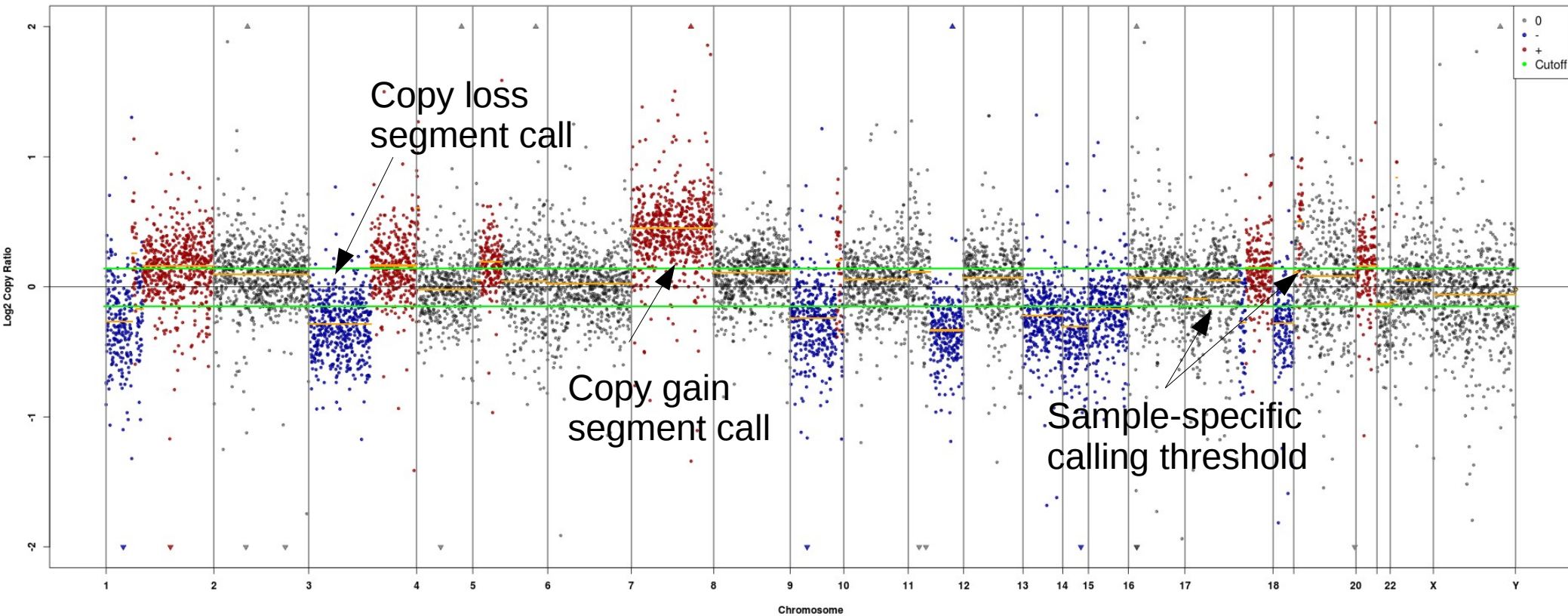
Normalization: Stage 2

Loess – remove residual GC bias

GC bias can still remain after normalization, especially in cases where systematic bias in the tumor sample is not well represented in the PON. This bias is explicitly removed from the log2 copy values by fitting a Loess model and subtracting the predicted values.



Calling gains and losses



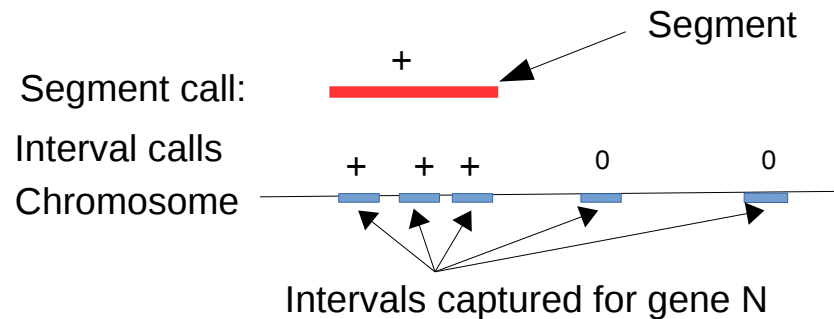
- Sample-specific calling threshold is a percentage* of the standard deviation of the intervals calculated as:

$$\text{median}([SD(\text{seg1 intervals}), SD(\text{seg2 intervals}) \dots SD(\text{segN intervals})]) * \text{tuning}$$

*The tuning parameter can be adjusted for higher sensitivity at the expense of specificity.

Segments → Intervals → Genes

- All callers produce segment calls
 - e.g. chr1:5000-20000 +
- Each baited interval is assigned a call based on the segment it intersects with.
- Gene calls are derived from the set of interval calls.



Gene Call: 3:+, 2:0 → Gain

Rules for Gene-level calls

<u>Rule</u>	<u>Gene-level Call</u>
'-' call > 2 times and '+' > 50%	'gain+loss'
'-' call > 2 or is 100%	'loss'
'+' call > 50%	'gain'
'+' and '-' calls in the same gene (below threshold)	'mixed'
'+' calls but below threshold	'Normal+'
'-' calls but below threshold	'Normal-'
No + or - calls	'Normal'

Outputs: Gene level calls

[illegible]

Outputs: Summarized gene level calls

Gene	gain	gain+loss	loss	mixed	NormalCopy	NormalCopy-	NormalCopy+	
PLCG1		28	0	1	0	57	0	0
CDK6		26	0	0	0	60	0	0
CHR7-18		26	0	0	0	60	0	0
CHR7-19		26	0	0	0	60	0	0
CHR7-20		26	0	0	0	60	0	0
chr7:8698365		26	0	0	0	60	0	0
HGF		26	0	0	0	60	0	0
HIP1		26	0	0	0	60	0	0
PTPN12		26	0	0	0	60	0	0
ABCB1		25	0	0	0	59	0	2
CARD11		25	0	0	0	61	0	0
CHR7-15		25	0	0	0	61	0	0
CHR7-16		25	0	0	0	61	0	0
CHR7-21		25	0	0	0	61	0	0

Outputs: QC Metrics

Sample	MaxNormalCor	MaxTumorCor	MeanCoverage	dQC	Var	GCscore	Cutoff	minPctTumorPurity1Gain	minPctTumorPurity1Loss
Sample1	0.652	0.698	34.4	0.174	0.275	3.8	0.252	38.2	32.1
Sample2	0.612	0.837	124	0.167	0.201	7.5	0.218	32.6	28
Sample3	0.608	0.835	122	0.12	0.142	4.6	0.151	22.1	19.9
Sample4	0.604	0.91	195	0.0349	0.756	9.5	0.37	58.5	45.2
Sample5	0.662	0.735	57.5	0.0924	0.127	1.9	0.161	23.6	21.1
Sample6	0.578	0.727	81.7	0.155	0.193	6.1	0.183	27	23.8
Sample7	0.522	0.83	120	0.154	0.155	6.4	0.187	27.7	24.3
Sample8	0.534	0.836	139	0.161	0.175	7.3	0.19	28.2	24.7
Sample9	0.49	0.775	106	0.139	0.276	7.1	0.197	29.3	25.5
Sample10	0.473	0.676	65.4	0.104	0.25	4	0.184	27.2	23.9
Sample11	0.545	0.701	76.3	0.0571	0.262	4.5	0.183	27	23.8
Sample12	0.546	0.837	95.6	0.111	0.268	8	0.24	36.2	30.7
Sample13	0.508	0.82	92.5	0.112	0.324	8.5	0.255	38.7	32.4

Minimum purity at which a single copy gain/loss could be detected.

