

# Gencove

Low coverage sequencing with imputation as an  
alternative traditional methods

# Gencove – Theory

Low-coverage sequencing

Imputation of genotypes based on algorithms and a dataset

## **Questions**

Does it call variants accurately across the genome

Is it a cost efficient alternative to standard sequencing of normal samples for tumor/normal analysis?

Can it be used instead of genotyping

# Method

6 dogs sequenced (~22X)

- 4 Flat-coated Retrievers
- 1 German shepherd
- 1 Cocker Spaniel

Low-coverage (~1X) fastq files extracted

Uploaded to Gencoves webpage and analyzed

# Method

## Output

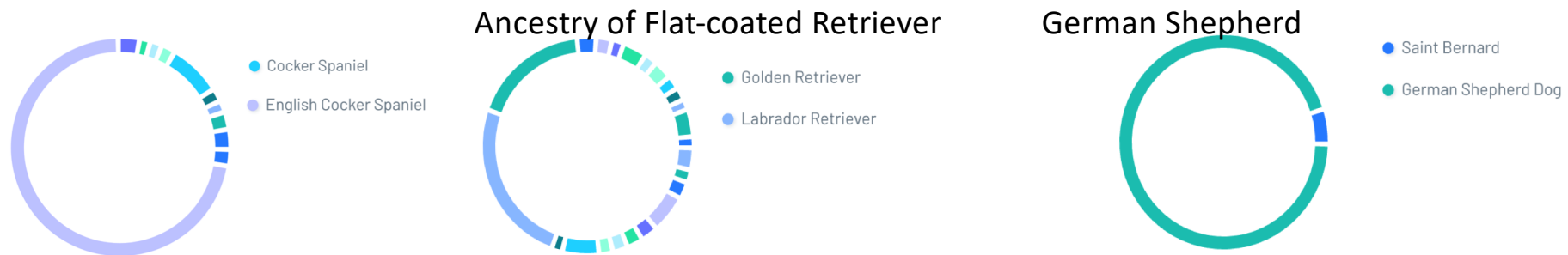
- Aligned sequence (BAM/BAI file)
- Ancestry (json)
- Microbiome
  - Unmapped sequences to the targeted genome are investigated for ancestry
- VCF files

# Breed analysis

The Flat-coateds are found to be mixes of primarily Labradors and Golden Retrievers

German Shepherd is 5% Saint Bernhard

Cocker Spaniel is 76% English Cocker Spaniel and 8% Cocker Spaniel



# Genotype Calling - Filters

## High coverage

- $QD < 2.0$
- $FS > 60.0$
- $SOR > 3.0$
- $MQ < 40.0$
- $MQRankSum < -12.5$
- $ReadPosRankSum < -8.0$
- $MAF \ 0.00001$

## Low coverage

- $GP > 0.9$

Or

- $GP > 0.7$

# Genotyping – Summary of Results – GP 0.9

No difference between breeds

All positions and all variants

- 378530 wrong variants
  - ~6 %
- 1848160 missed variants
  - ~29 %

All positions and only SNPs

- 47063 wrong calls
  - ~1 %
- 11941456 missed calls
  - ~26 %

174k positions (only SNPs)

- 183 wrong calls
  - 0.003%
- 3735 missed calls
  - ~4%

# Genotyping – Summary of Results – GP 0.7

No difference between breeds

All positions and all variants

- 486698 wrong variants
  - ~8 %
- 1726881 missed variants
  - ~27%

All positions and only SNPs

- 62065 wrong calls
  - ~1 %
- 1158178 missed calls
  - ~25 %

174k positions (only SNPs)

- 233 wrong calls
  - 0.3 %
- 3092 missed calls
  - ~4 %



# Comparison

If lowering min(GP) from 0.9 to 0.7

All positions and all variants

- Wrong variants:
  - 6 % -> 8 %
- Missed variants:
  - 29 % -> 27 %

All positions and only SNPs

- Wrong calls
  - 1 % -> 1 %
- Missed calls
  - 26 % -> 25 %

174k positions and only SNPs

- Wrong calls
  - 0.2 % -> 0.3 %
- Missed calls
  - 5 % -> 4 %

# Comparisons – All variants – GP 0.9

174 k variants	Mean		FCR1		FCR2		FCR3		FCR4		Cocker		German	
Complete matches	80779	95%	80641	95%	79873	95%	79062	95%	80356	96%	82460	95%	82282	96%
Missed variants	3818	5%	4135	5%	3950	5%	3984	5%	3400	4%	3764	4%	3672	4%
Variants called, but wrong	5.5	0%	4	0%	6	0%	2	0%	7	0%	7	0%	7	0%
Wrong calls	185	0%	150	0%	152	0%	202	0%	156	0%	285	0%	165	0%
All mistakes	191	0.2%	154	0%	158	0%	204	0%	163	0%	292	0%	172	0%
Total	84787	100%	84930	100%	83981	100%	83250	100%	83919	100%	86516	100%	86126	100%
All positions	Mean		FCR1		FCR2		FCR3		FCR4		Cocker		German	
Complete matches	4084399	65%	4076664	67%	4042884	63%	3984030	63%	4063340	64%	4159387	68%	4180087	63%
Missed variants	1848160	29%	1630666	27%	2029976	31%	1915865	31%	1902953	30%	1542112	25%	2067390	31%
Variants called, but wrong	258484	4%	261621	4%	261723	4%	251548	4%	272155	4%	266118	4%	237738	4%
Wrong calls	120047	2%	115344	2%	112198	2%	122644	2%	123006	2%	137163	2%	109927	2%
All mistakes	378531	6%	376965	6%	373921	6%	374192	6%	395161	6%	403281	7%	347665	5%
Total	6311090	100%	6084295	100%	6446781	100%	6274087	100%	6361454	100%	6104780	100%	6595142	100%

# Comparisons – All variants – GP 0.7

174 k variants	Mean		FCR1		FCR2		FCR3		FCR4		Cocker		German	
Complete matches	81423.8	96%	81490	96%	80639	96%	79826	96%	80767	96%	83038	96%	82783	96%
Missed variants	3170	4%	3284	4%	3180	4%	3215	4%	2987	4%	3183	4%	3171	4%
Variants called, but wrong	8.16667	0%	6	0%	10	0%	7	0%	9	0%	10	0%	7	0%
Wrong calls	236	0%	206	0%	201	0%	264	0%	194	0%	340	0%	211	0%
All mistakes	244	0.3%	212	0%	211	0%	271	0%	203	0%	350	0%	218	0%
Total	84838	100%	84986	100%	84030	100%	83312	100%	83957	100%	86571	100%	86172	100%
All positions	Mean		FCR1		FCR2		FCR3		FCR4		Cocker		German	
Complete matches	4153771	65%	4152608	68%	4117894	63%	4059473	64%	4117029	64%	4223754	69%	4251869	64%
Missed variants	1726881	27%	1501732	24%	1902305	29%	1788060	28%	1800581	28%	1429649	23%	1938958	29%
Variants called, but wrong	265419	4%	314611	5%	314384	5%	303910	5%	178112	3%	314214	5%	167284	3%
Wrong calls	221279	3%	171350	3%	166785	3%	182724	3%	320838	5%	191587	3%	294388	4%
All mistakes	486698	8%	485961	8%	6501368	100%	486634	8%	498950	8%	505801	8%	461672	7%
Total	6367350	100%	6140301	100%	6501368	100%	6334167	100%	6416560	100%	6159204	100%	6652499	100%

# Comparisons – SNPs only – GP 0.9

174 k variants	Mean		FCR1		FCR2		FCR3		FCR4		Cocker		German	
Complete matches	80745	96%	80608	95%	79839	95%	79030	95%	80319	96%	82427	96%	82247	96%
Missed variants	3735	4%	4052	5%	3868	5%	3901	5%	3324	4%	3677	4%	3589	4%
Variants called, but wrong	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%
Wrong calls	183	0.003%	147	0%	150	0%	201	0%	154	0%	281	0%	163	0%
All mistakes	183	0.003%	147	0%	150	0%	201	0%	154	0%	281	0%	163	0%
Total	84480	100%	84660	100%	83707	100%	82931	100%	83643	100%	86104	100%	85836	100%
All positions	Mean		FCR1		FCR2		FCR3		FCR4		Cocker		German	
Complete matches	3360320	73%	3351113	76%	3320982	70%	3277988	71%	3335116	72%	3422378	78%	3454340	71%
Missed variants	1194146	26%	987117	23%	1355643	29%	1280121	28%	1270050	27%	927973	21%	1343970	28%
Variants called, but wrong	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%
Wrong calls	47063	1%	44226	1%	43734	1%	48149	1%	45909	1%	53407	1%	46953	1%
All mistakes	47063	1%	44226	1%	43734	1%	48149	1%	45909	1%	53407	1%	46953	1%
Total	4601528	100%	4382456	100%	4720359	100%	4606258	100%	4651075	100%	4403758	100%	4845263	100%

# Comparisons – SNPs only – GP 0.7

174 k variants	Mean		FCR1		FCR2		FCR3		FCR4		Cocker		German	
Complete matches	81388.2	96%	81453	96%	80604	96%	79792	96%	80730	96%	83004	96%	82746	96%
Missed variants	3092	4%	3207	4%	3103	4%	3139	4%	2913	3%	3100	4%	3090	4%
Variants called, but wrong	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%
Wrong calls	233	0.275%	202	0%	199	0%	262	0%	191	0%	335	0%	209	0%
All mistakes	233	0.275%	202	0%	199	0%	262	0%	191	0%	335	0%	209	0%
Total	84713	100%	84862	100%	83906	100%	83193	100%	83834	100%	86439	100%	86045	100%
All positions	Mean		FCR1		FCR2		FCR3		FCR4		Cocker		German	
Complete matches	3396288	74%	3392877	77%	3361892	71%	3319075	72%	3359889	72%	3455501	78%	3488491	72%
Missed variants	1158178	25%	945353	21%	1314733	28%	1239034	27%	1245277	27%	894850	20%	1309819	27%
Variants called, but wrong	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%
Wrong calls	62065.2	1%	59729	1%	58693	1%	64367	1%	59219	1%	68139	2%	62244	1%
All mistakes	62065.2	1%	59729	1%	58693	1%	64367	1%	59219	1%	68139	2%	62244	1%
Total	4616530	100%	4397959	100%	4735318	100%	4622476	100%	4664385	100%	4418490	100%	4860554	100%

# Conclusions

- Very low %-percentage are called wrong
- Still around 200 wrong calls
  - Will it change results of analysis?
- Lowering min(GP) lead to more mistakes but also less missed variants
- Suitable for genotyping?
- Not sufficiently precise for tumor-normal data