GENE EDITING IN THE CONTEXT OF GENOME VARIATION

Elena Rice

CSO

GENOME INFORMATION AVAILABLE IN PUBLIC DATABASES

Genome Information	Human (GRCh38.p14)	Mouse (GRCm39)	Cow (ARS-UCD1.3)	Pig (Sscrofa11.1)
Base pairs	3,099,750,718	2,728,222,451	2,711,209,831	2,501,912,388
Coding genes	23,132	21,955	23,842	22,063
Non-coding genes	28,471	18,008	11,707	13,154
Short variants	1,110,229,688	122,449,497	96,494,428	71,053,233
Structural variants	7,861,655	791,878	18,942	224,038

- Genome sizes are very similar
- **Similar number** of coding genes across species
- **Significantly different number** of short and structural variants compared to Human alternative assemblies
- Pan genome efforts in cows and pigs will expand our genomic understanding beyond single breed/animals

There is more variation to capture

Only one reference genome assembly - Sscrofa11.1 plus 12 additional breeds

Ensembl variant database: total SNPs: 71,053,233

European Variant Archive: total SNPs 72,607,645

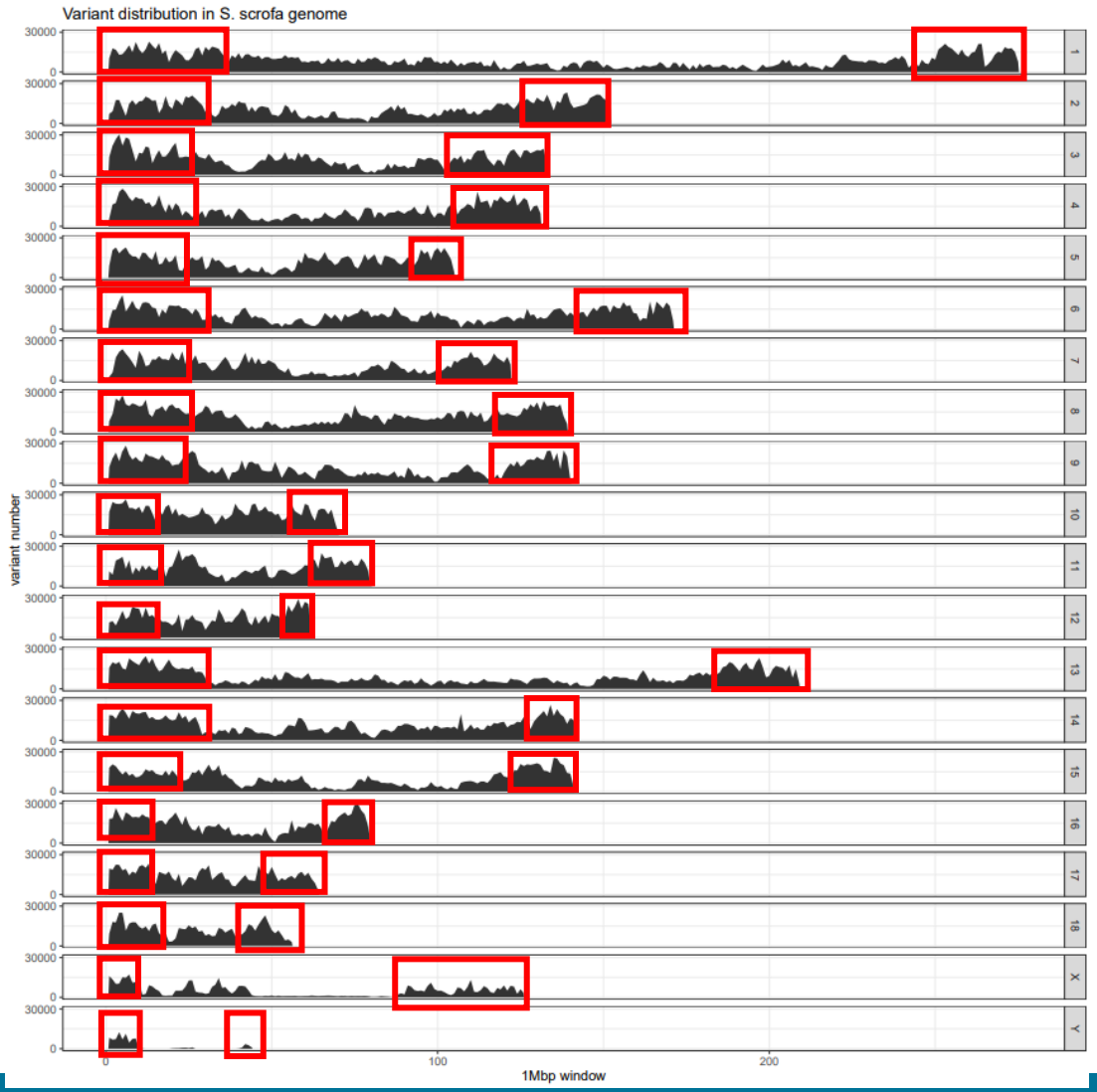
Genus Whole Genome Sequence data (7000 samples): Total SNPs: 101,012,026

SNPs not present in public databases: 37,021,234

Natural variation across individuals/breeds is not evenly distributed in the pigs' genome

- 2/3 of the variants are in highly repetitive sequence domains in telomeric regions
- **A small proportion of variants are in coding genes**

NATURAL VARIATION ACROSS INDIVIDUALS/BREEDS IS NOT EVENLY DISTRIBUTED IN THE PIGS' GENOME



Ideogram showing the distribution of natural variants in swine chromosomes

- Two-thirds of the variants are in highly repetitive sequence domains in telomeric regions
- **A small proportion of variants are in coding genes**
- Functional annotation provides information about the impact on coding genes

SNP effects based on gene functional annotation in coding regions:

stop gain/lost: 8,092
synonymous: 287,606
missense: 227,894

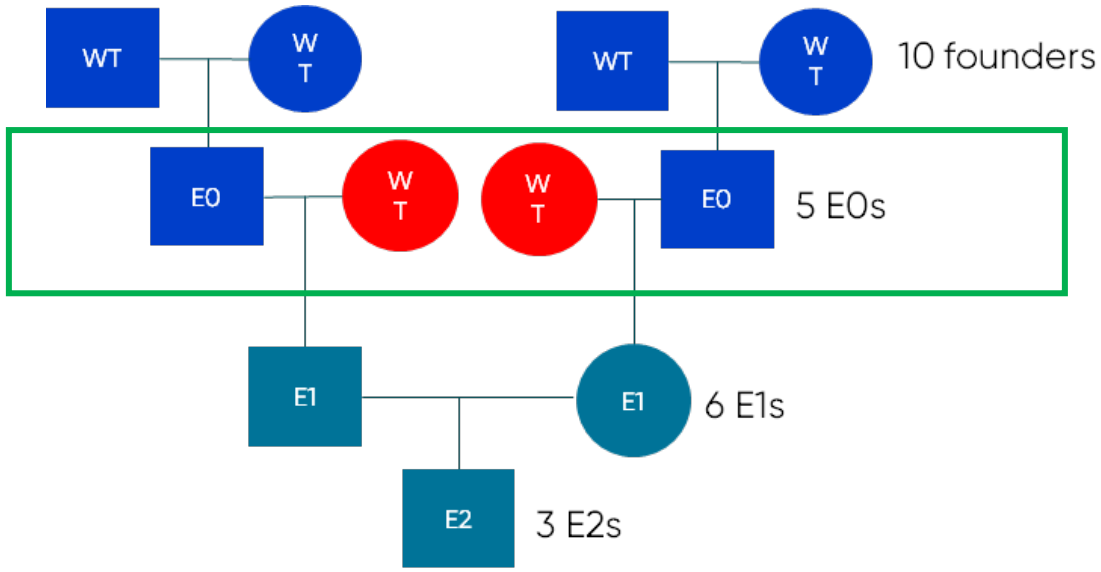
SNPs not present in public databases in coding regions

stop gain/lost: 14,394
synonymous: 284,502
missense: 217,793

Genus

DE NOVO NATURAL VARIANTS VS OFF-TARGET EVENTS

***De novo* vs Off-target variants analysis on E0 pigs using WGS (short and structural variants from reference swine population)**

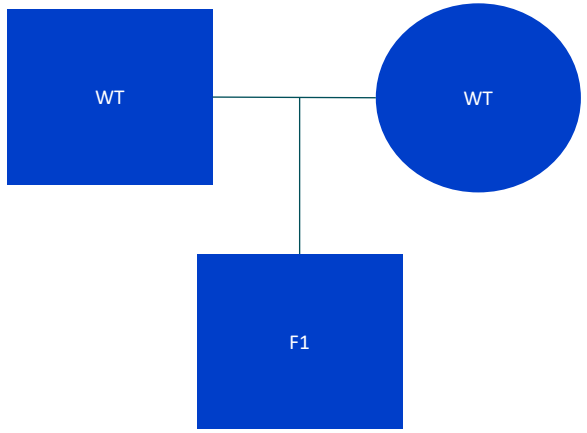


- Total of 1225 *de novo* variants across 5 E0s were found using existing reference genome SNP variation
- **192-466 p**otential “de novo” mutations per individual
- Not possible to differentiate between true *de novo* (not present in the parental samples) vs simply not represented in the reference population
- Using guides’ sequences can identify small guide-specific changes outside of the target edit
- What do the rest of the “new” variants represent?
  - Large number of false positives - these variants are not truly *de novo* they are just **not represented in the population database**
  - We can confirm that in fact, they are inherited from one of the parents

EVALUATION OF LEVEL OF NATURALLY OCCURRING *DE NOVO* **VARIANTS** IN PIGS THROUGH **NORMAL REPRODUCTION CYCLES**

De novo variant - a genetic change (SNP or INDEL) that is present in the progeny and absent in the parental samples

Whole Genome Sequencing of 13 litters in trios



De novo variants defined as:

- Not present in either parent
- Heterozygous (0/1) in F1 offspring
- Criteria used: allele frequency >0.2 with depth of alternative allele coverage ≥20X coverage

F1 animals

Total *de novo* variants - 1367
88-134 variants per individual vs 192-466

Total Indels - 394
24-43 indels per individual

Total *de novo* variants in coding regions - 37

Missense - 2

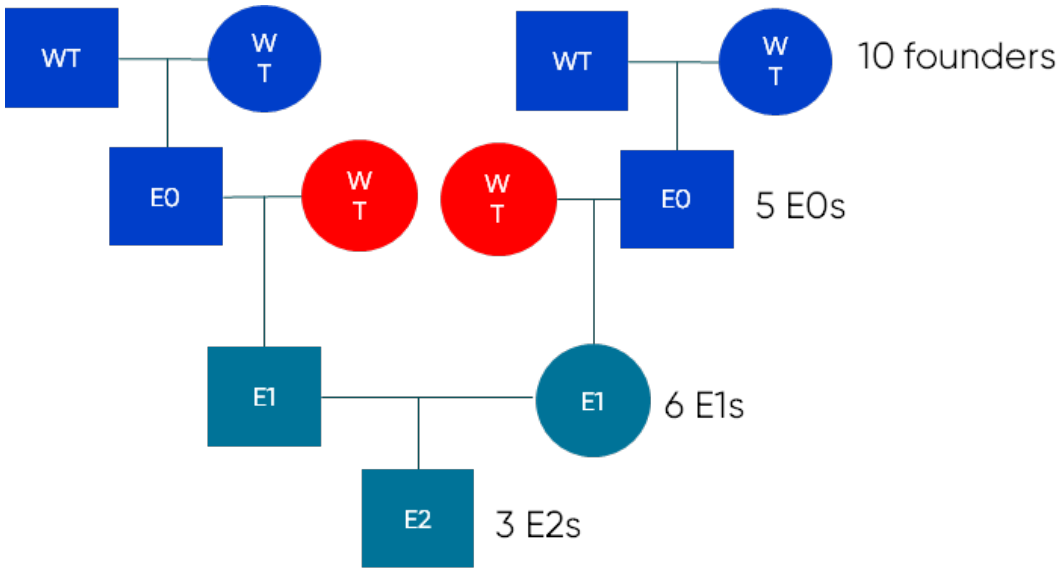
Only the family trio analysis allowed us to detect true de novo events in each generation

Only a small number of de novo variation is in coding regions

Genus

DE NOVO VARIANTS IN GENE-EDITED PIGS DEMONSTRATE THAT **OFF-TARGETS ARE NOT DRIVEN BY RANDOM EVENTS IN THE GENOME**

Whole Genome Sequencing in trios



- Trio analysis was done for 24 sequenced gene-edited animals
- 2 TB of processed data was produced
- Sequencing depth: 30X to 60X
- Cost: average $1,500 per sample depending on depth

Identified true *de novo* variants in gene-edited pigs:

E0= 63 (SD=8)+17 off-targets
E2= 80 (SD=15) – no off-targets

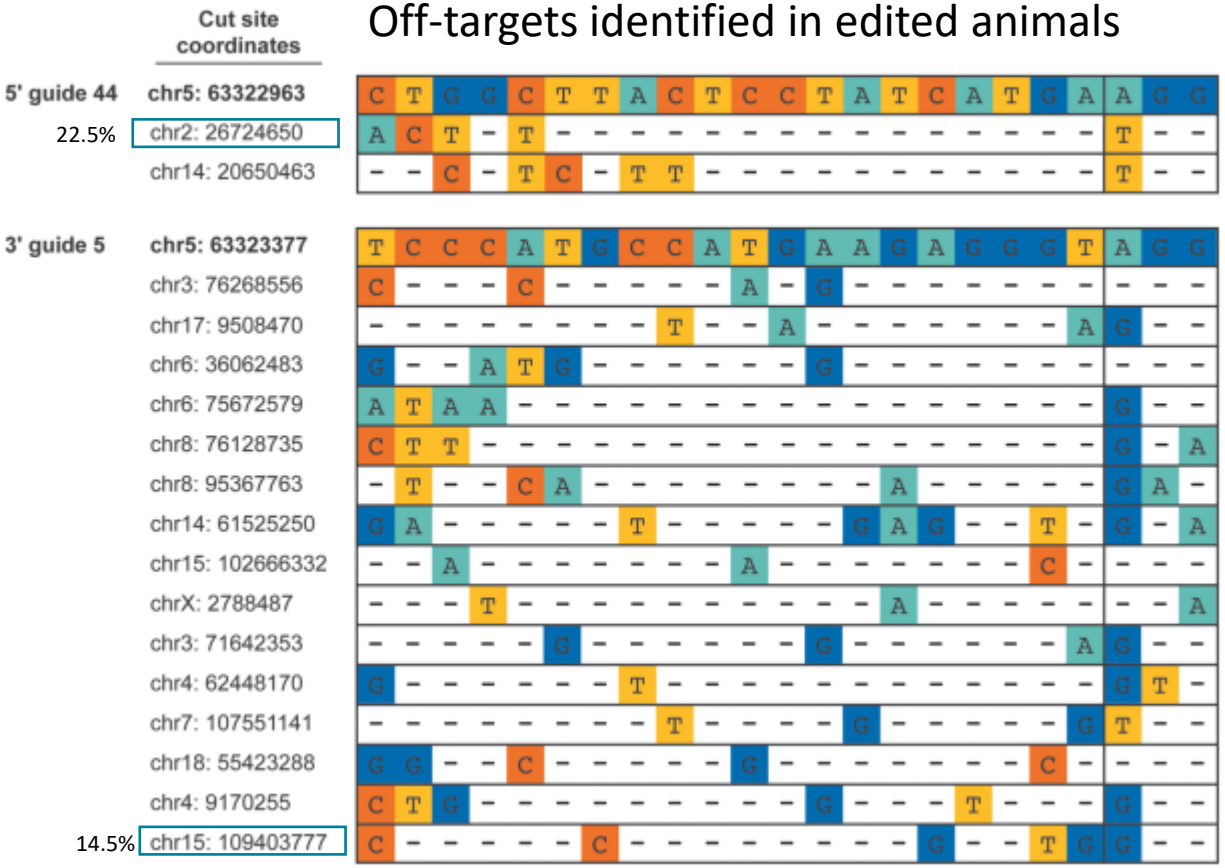
- **None *de novo* variants are in target regions or genes**
- **No significant difference in frequency of *de novo* variant in edited vs non-edited pigs**
- **Off-target events in gene-edited pigs are not random (guide and method driven)**

IN THE CONTEXT OF GENE EDITING: TARGETED APPROACH ALLOWS FOR PRECISE OFF-TARGET IDENTIFICATION. WGS ANALYSIS CONFIRMS THAT OFF-TARGETS ARE NOT DRIVEN BY RANDOM EVENTS IN GENOME

Analysis of off-targets for PRRSV-resistant pigs

- Site–seq – identifies **potential** off-targets based on guide sequences:
  - 182 potential off-targets (in silico and in vitro analysis)
- Sequence capture – confirms **actual** off-targets in edited animals (E0)
  - Only 17 off-targets found in edited animals
- PCR screening – to reconfirm off-targets in edited animals
  - Primers for 5’ and 3’ guides
  - All E0 tested with negative results

Off-targets identified in edited animals



Top of each panel shows Cas9 spacer sequence and PAM for 5’ and 3’ guides. Conserved nucleotides are depicted with a dash

SUMMARY

- There is more diversity in animal genomes then captured today. To capture and represent true individual diversity, Alternative Genome Assemblies, Pan-genomes, and precise annotation is needed
- Variation in the genome is not random (hot spots)
- Off-target events in gene-edited pigs are not random (guides and method driven)
- Off-target edits are reliably detected via targeted analysis of predicted off-target sites. There are in silico, in vitro, and in vivo methods to test off-target events that can be combined to minimize false positive and false negative results
- To use WGS for off-target identification, the generational trios should be used at high sequencing depth (60X-100X). The same off-targets were identified by targeted analysis. WGS adds significant cost, time, and computing resources and, therefore, not practical
- Naturally occurring de novo events are present in all individuals in each generation (63-134). The family trio analysis was the most accurate way to capture those events in pig populations
- CRISPR/Cas-edited pigs showed a lower rate of off-target mutation than natural variation in conventionally bred pigs