Identification of single-nucleotide variants associated with susceptibility to *Salmonella* in pigs using a genome-wide association approach

Corinne H. Schut1, Abdolvahab Farzan1,2, Russell S. Fraser1,3, Margaret H. Ainslie-Garcia1, Robert M. Friendship2 and Brandon N. Lillie1*

**Abstract**

**Background:** *Salmonella enterica* serovars are a major cause of foodborne illness and have a substantial impact on global human health. In Canada, *Salmonella* is commonly found on swine farms and the increasing concern about drug use and antimicrobial resistance associated with *Salmonella* has promoted research into alternative control methods, including selecting for pig genotypes associated with resistance to *Salmonella*. The objective of this study was to identify single-nucleotide variants in the pig genome associated with *Salmonella* susceptibility using a genome-wide association approach. Repeated blood and fecal samples were collected from 809 pigs in 14 groups on farms and tonsils and lymph nodes were collected at slaughter. Sera were analyzed for *Salmonella* IgG antibodies by ELISA and feces and tissues were cultured for *Salmonella*. Pig DNA was genotyped using a custom 54 K single-nucleotide variant oligo array and logistic mixed-models used to identify SNVs associated with IgG seropositivity, shedding, and tissue colonization.

**Results:** Variants in/near *PTPRJ* (p = 0.0000066), *ST6GALNAC3* (p = 0.0000099), and *DCDC2C* (n = 3, p < 0.0000086) were associated with susceptibility to *Salmonella*, while variants near *AKAP12* (n = 3, p < 0.0000358) and in *RALGAPA2* (p = 0.0000760) may be associated with susceptibility.

**Conclusions:** Further study of the variants and genes identified may improve our understanding of neutrophil recruitment, intracellular killing of bacteria, and/or susceptibility to *Salmonella* and may help future efforts to reduce *Salmonella* on-farm through genetic approaches.

**Keywords:** *Salmonella*, Swine, Antibody response, Shedding, GWAS, Single-nucleotide variant
Background
Salmonella is one of the leading causes of foodborne illness and has a significant impact on human health both globally and in Canada [1–3]. While eggs and poultry are the most frequently identified sources of human salmonellosis, pork is also a notable source of Salmonella [4–7]. Studies assessing Salmonella prevalence through serology and/or culture have frequently identified Salmonella in pigs in North America and Europe [8–12]. In pigs, S. Choleraesuis infection typically manifests as swine typhoid that may result in diarrhea, fever, and septicemia, similar to human-infecting typhoidal Salmonella serovars like S. Typhi [13]. Pigs showing visible signs of illness may be treated or removed from the herd to reduce the spread of Salmonella. However, the most frequently identified serovars on Canadian swine farms included S. Typhimurium, S. Typhimurium var. Copenhagen, and S. Infantis [9, 14, 15] which typically result in an asymptomatic carrier state in pigs but are known to cause illness in humans [16]. Pigs carrying Salmonella asymptomatically play a significant role in on-farm transmission of Salmonella within the herd and may limit the effectiveness of control measures implemented on-farm [12]. On-farm control of Salmonella has consisted of stringent biosecurity and sanitation practices, as well as the use of antibiotics, vaccination, and quarantine or culling of infected pigs [17–20]. However, the limited effectiveness of these measures in practice has prompted research into swine genetics as a potential alternative measure to control Salmonella on swine farms.

Traditionally, selective breeding in swine was established to promote desired production traits including growth performance, feed efficiency, fertility, and meat quality [21–23]. However, with the completion and continued updates to the porcine genome, many studies are now investigating the genetic basis of disease susceptibility in swine. One approach in using genetics to improve resistance is to observe immune traits or phenotypes individually (for example; cytokine production, leukocyte proliferation, and serum levels of IgG or acute phase proteins) [24–27]. Differences in these immune traits and disease severity between pigs and between breeding lines has been well documented which suggests the potential of selective breeding for improved resistance in the near future [19, 24, 28–30]. One such study found that piglets with improved recruitment and function of polymorphonuclear neutrophils, but a lower antibody response, were more resistant to Salmonella [28]. As such, it may be possible to select from these breeding lines with more robust immune response phenotypes or desired response traits to promote broad immunity to Salmonella in offspring.

Beyond the assessment of immune traits, several studies in recent years have identified significant associations between single-nucleotide variants (SNVs) and/or candidate genes and susceptibility to Salmonella in pigs. Candidate gene studies have observed variants in porcine toll-like receptor (TLR) genes that were associated with Salmonella fecal shedding [31], and attenuated responses to Salmonella Choleraesuis [32]. Upregulation of TLR5 and TLR9 has been shown in response to S. Choleraesuis and S. Typhimurium though its direct impact on Salmonella susceptibility is unknown [33]. Additionally, SNVs in mannann-binding lectin (MBL) 1 and 2, have been found at higher frequencies in pig populations infected with S. Typhimurium and other pathogens [34]. Further, a candidate gene study of the pigs included in this study identified an MBL1 variant associated with increased Salmonella shedding and a variant in NOD1 associated with isolation of Salmonella at slaughter [35].

The candidate gene studies may offer insight into pig susceptibility to Salmonella on-farm and at slaughter and benefit efforts in breeding for resistance to common pathogens on-farm. However, a major drawback of candidate gene studies is that they require a priori knowledge of these genes and their functions, and there is still much that is unknown about the pig immune response and the complex interplay between pathogen and host [36]. With recent technological advancements improving the feasibility of genome-wide association studies (GWAS), we can potentially identify novel variants associated with resistance to Salmonella shedding and colonization across the entire genome [37]. This study aimed to identify SNVs associated with Salmonella IgG antibody response from the end of nursery to market, Salmonella shedding from weaning to market, and presence of Salmonella in tonsil and lymph node tissues at slaughter in commercial pigs using a GWAS approach.

Results
Of the pigs included in the GWAS seropositivity model, 32.3% (254/786) of pigs were seropositive at least once from the end of the nursery stage to the end of the finisher stage, for the shedding model 34.2% (269/786) of pigs shed at least once from weaning to end of finisher, and for colonization model 21.6% (111/515) of pigs tested positive for Salmonella at slaughter. Salmonella positivity for each trait is shown in Table 1.

Genome-wide association study
After quality control filtering of the data, 51,969 SNVs remained for GWAS analysis from 786 pigs in the seropositivity and shedding models and 515 pigs in the slaughter model. Manhattan plots and quantile-quantile plots for the seropositivity, shedding, and colonization phenotypes are shown in Figs. 1, and 2, respectively. The
allele frequencies between case and control pigs for the five significant SNVs and SNVs approaching significance in the GWAS analysis can be seen in the Supplementary Table 1.

Analysis of the shedding trait identified an intron variant in the gene PTPRJ on chromosome 2 \((p = 6.6 \times 10^{-6})\) as well as an intergenic variant on chromosome 6, downstream of the ST6GALNAC3 gene \((p = 9.9 \times 10^{-6})\), that had a suggestive association (Table 2) and were observed at a higher frequency in the pigs that did not shed Salmonella. GWAS analysis of Salmonella isolation at slaughter indicated three intergenic SNVs downstream of the gene DCDC2C on chromosome 3 \((p = 1.6 \times 10^{-6}, 3.7 \times 10^{-6}, 8.6 \times 10^{-6}, \text{respectively})\) under the suggested association threshold of \(5.0 \times 10^{-5}\) (Table 3), two of which were seen at higher frequency in pigs that tested negative at slaughter, while one was more common in pigs positive at slaughter. Three intergenic SNVs on chromosome 1, upstream of the gene AKAPI2 \((p = 2.0 \times 10^{-5}, 3.1 \times 10^{-5}, 3.6 \times 10^{-5}, \text{respectively})\), also approached the suggested association threshold and were more common in the pigs that were negative at slaughter. An intron SNV located in AKAPI2 was also observed in the top 15 SNVs for the slaughter model (Table 3). No SNVs were under the suggestive association threshold of \(5.0 \times 10^{-5}\) for the analysis.

### Table 1 Demographics in pigs positive and negative for the trait of interest after quality control of genotypic data

| Phenotype                        | Number of negative pigs (controls*) | Number of positive pigs (cases*) |
|----------------------------------|-------------------------------------|----------------------------------|
| Salmonella seropositivity        | 254 (32.3%)                         | 532 (67.7%)                      |
| Salmonella shedding              | 269 (34.2%)                         | 517 (65.8%)                      |
| Salmonella isolation at slaughter| 111 (21.6%)                         | 404 (78.4%)                      |

*Control: never tested seropositive, never shed Salmonella, or tested negative at slaughter; Case: tested seropositive for Salmonella at least once, shed Salmonella at least once on-farm, or was positive for Salmonella at slaughter

---

**Fig. 1** Manhattan plots of the GWAS analysis for Salmonella seropositivity from end of nursery to end of finisher \((a)\), Salmonella shedding from weaning to end of finisher \((b)\), and Salmonella isolation from tissues at slaughter \((c)\). The horizontal solid and dashed red lines indicate the genome-wide threshold for significant \((p = 5.0 \times 10^{-7})\) and suggestive \((p = 1.0 \times 10^{-5})\) associations, respectively [38].
of the seropositivity trait. An intron variant located within the \textit{RALGAPA2} gene on chromosome 17 was the most significant SNV associated with a pig being seropositive at least once from nursery to finisher stage ($p = 7.6 \times 10^{-5}$). Additionally, three intron SNVs in the \textit{RALGAPA2} gene were observed within the top 15 significant SNVs for the seropositivity model (Table 4).

\section*{Discussion}

This study aimed to assess susceptibility to \textit{Salmonella} in commercial pigs by using a GWAS approach to identify potential variants associated with \textit{Salmonella} IgG antibody response from end of nursery to end of finisher, \textit{Salmonella} shedding from weaning to end of finisher, and isolation of \textit{Salmonella} from tissues at slaughter. GWAS is a useful tool in preliminary identification of novel SNVs or genes and associations between genetic variants and disease resistance or susceptibility. This has promise for future efforts in breeding for resistance to \textit{Salmonella} on-farm and at slaughter, which may lead to a reduction of \textit{Salmonella} benefiting both public health and animal welfare.

\textbf{GWAS analysis of shedding from nursery to finisher stage}

GWAS analysis identified two associated SNVs. The first was an intron variant located in the \textit{PTPRJ} gene, which encodes protein tyrosine phosphatase (PTP) receptor type J (PTPRJ) and was observed at a higher frequency in pigs that never shed \textit{Salmonella}. These PTPRJ receptors are found on many cell types and are highly expressed in the macrophage rich tissues of the intestines [40]. PTPRJ receptors are responsible for negative regulation of protein tyrosine kinases, which are involved in pathogen recognition and clearance, as well as resolution of inflammation [41–44]. PTPs may also be involved in B cell activation, recruitment of neutrophils, and the negative regulation of T-cell signaling [39, 45–47]. The timely recruitment of neutrophils is a crucial part of the immune response and often plays a key role in the clearance of pathogens in the gut [48]. A challenge study on \textit{PTPRJ}-deficient mice showed impairment...
Table 2 The top 15 SNVs ranked by significance (p value) for the GWAS analysis of Salmonella shedding from weaning to end of finisher

| SNV ID       | SSC  | Location (bp) | Variant type | Gene       | Gene location | p-value       |
|--------------|------|---------------|--------------|------------|---------------|---------------|
| rs324041697  | 2    | 14,596,159    | intron       | PTPRJ      | 2: 14549537–14,726,715 | 6.6 × 10^{-6} |
| rs81476180   | 6    | 136,607,113   | intergenic   | ST6GALNAC3 | 6: 136633437–137,201,213 | 9.9 × 10^{-6} |
| rs81259485   | 16   | 3,248,143     | intron       | DNAH5      | 16: 3116189–3,364,956  | 6.6 × 10^{-5} |
| rs81458700   | 16   | 36,370,940    | intergenic   | ACTBL2     | 16: 3647053–36,471,633 | 9.9 × 10^{-5} |
| rs80889714   | 4    | 104,920,198   | intron       | CASQ2      | 4: 104918501–104,982,852 | 1.0 × 10^{-4} |
| rs81346930   | 17   | 44,032,145    | intergenic   | LPIN3      | 17: 43993710–44,006,188 | 1.2 × 10^{-4} |
| rs330624049  | 15   | 90,476,754    | intergenic   | ENSSSCG0000034554 | 15: 90309385–90,309,702 | 1.2 × 10^{-4} |
| rs81403742   | 8    | 117,467,555   | intergenic   | TACR3      | 8: 117661792–117,721,591 | 1.2 × 10^{-4} |
| rs80947769   | 7    | 19,456,263    | upstream     | KIAA0319   | 7: 19386111–19,451,666  | 1.2 × 10^{-4} |
| rs339478050  | 1    | 188,917,080   | intron       | RTN1       | 1: 188704139–188,926,054 | 1.9 × 10^{-4} |
| rs81237965   | 3    | 3,693,136     | intron       | RADIL      | 3: 3618093–313,033    | 1.9 × 10^{-4} |
| rs81344023   | 6    | 153,022,279   | intron       | FGGY       | 6: 152754730–153,213,125 | 2.1 × 10^{-4} |
| rs81349902   | 1    | 188,903,236   | intron       | RTN1       | 1: 188704139–188,926,054 | 2.3 × 10^{-4} |
| rs318950111  | 3    | 43,561,730    | downstream   | ZNF169     | 3: 43513397–43,558,613 | 2.5 × 10^{-4} |
| rs81452423   | 15   | 32,935,177    | intergenic   | DLGAP2     | 15: 33173120–33,264,155 | 3.0 × 10^{-4} |

SSC = Sus scrofa chromosome
Location in Ensembl Suscrofa11.1
If the variant was intergenic, the closest gene within a 1 Mbp window was indicated
SNVs under the suggestive significance threshold of 1.0 × 10^{-5} in bold [39]

of B cell development but an unexpected significant increase in early neutrophil recruitment and rapid clearance of Staphylococcus aureus in a subcutaneous air pouch and suggested that PTPRJ plays both a positive and negative role in attraction of neutrophils [49].

Interestingly, the PTPRJ variant was observed at a significantly higher frequency in pigs with no Salmonella shedding observed from weaning to end of finisher. It is possible that, this variant may be causing a reduction in the expression of PTPRJ resulting in a subsequent rapid

Table 3 Top 15 SNVs ranked by significance for the GWAS analysis of isolation of Salmonella from tissues at slaughter

| SNV ID       | SSC  | Location (bp) | Variant type | Gene       | Gene location | p-value       |
|--------------|------|---------------|--------------|------------|---------------|---------------|
| rs322440805  | 3    | 130,637,256   | intergenic   | DCDC2C     | 3: 131097730–131,194,726 | 1.6 × 10^{-6} |
| rs32641709   | 3    | 130,676,894   | intergenic   | DCDC2C     | 3: 131097730–131,194,726 | 3.7 × 10^{-6} |
| rs319944764  | 3    | 130,689,632   | intergenic   | DCDC2C     | 3: 131097730–131,194,726 | 8.6 × 10^{-6} |
| rs81348815   | 1    | 15,081,855    | intergenic   | AKAP12     | 1: 14906823–15,016,841 | 2.0 × 10^{-5} |
| rs80951933   | 1    | 15,109,785    | intergenic   | AKAP12     | 1: 14906823–15,016,841 | 3.1 × 10^{-5} |
| rs8093645    | 1    | 15,064,589    | intergenic   | AKAP12     | 1: 14906823–15,016,841 | 3.6 × 10^{-5} |
| rs326617356  | 18   | 26,691,633    | intergenic   | KCND2      | 18: 26554190–26,623,034 | 1.3 × 10^{-4} |
| rs323563819  | 19   | 109,560,455   | intron       | GPIC3      | X: 109536447–110,060,245 | 2.0 × 10^{-4} |
| rs80820138   | 14   | 122,383,711   | intergenic   | GPAM       | 14: 122542205–122,577,239 | 2.0 × 10^{-4} |
| rs80961723   | 7    | 120,400,325   | intergenic   | SETD3      | 7: 120412549–120,486,627 | 2.2 × 10^{-4} |
| rs320610499  | 6    | 141,349,179   | intron       | NEGR1      | 6: 141115729–141,647,820 | 2.5 × 10^{-4} |
| rs8126290    | 2    | 91,504,717    | intron       | VCAN       | 2: 91287931–91,811,674  | 2.8 × 10^{-4} |
| rs324243793  | 18   | 38,747,572    | intron       | DPY19L2    | 18: 38693003–38,764,024 | 3.3 × 10^{-4} |
| rs80846097   | 1    | 14,988,130    | intron       | AKAP12     | 1: 14906823–15,016,841 | 4.1 × 10^{-4} |
| rs81361262   | 2    | 90,665,766    | downstream   | ATG10      | 2: 90437020–90,665,380  | 4.5 × 10^{-4} |

SSC = Sus scrofa chromosome
Location in Ensembl Suscrofa11.1
If the variant was intergenic, the closest gene within a 1 Mbp window was indicated
SNVs under the suggestive significance threshold of 1.0 × 10^{-5} in bold [39]
increase in neutrophil recruitment and clearance of *Salmonella* infection in the control population [49], while RPTP, another member of the PTP family, is rescuing some B cell functionality and signalling [48]. Currently, more research is needed into the putative function of PTPR in pigs and how this variant and its interactions with other PTPs, like RPTP, may affect B cell signalling and neutrophil recruitment in response to pathogens like *Salmonella* before a definitive conclusion can be made.

Analysis of the shedding trait also identified an intergenic variant nearest to the gene encoding *ST6GALNAC3* which encodes ST6 (α-2,6-sialyltransferase 3) and was also more frequent in pigs that were not observed to be shedding from weaning to finisher stage. Due to the role of sialic acids in cell adhesion or nutrient acquisition and thus affecting either the availability and localization of host-derived sialylated glycans and impairing the rate of *Salmonella* adherence or nutrient acquisition and thus may affect *Salmonella* susceptibility in pigs.

**GWAS analysis of colonization at slaughter**

The model for isolation of *Salmonella* from tissues at slaughter identified three intergenic variants with a suggestive association nearest to the gene encoding double-cortin domain containing 2C (DCDC2C). Two of these variants were more frequent in pigs that tested negative at slaughter, while the third was seen at a higher frequency in pigs positive at slaughter. The *DCDC2C* gene is currently uncharacterized in pigs and little is known about its function in vivo. Studies of *DCDC2C* in
humans have suggested that variants in this gene may be associated with structural defects in cilia in sperm and in cilia length in sensory cells in the ear [60, 61]. Thus, it is possible that the variants observed in the current study also result in structural defects, either in the cilia in the lungs, or perhaps the villi of the intestines of pigs and may be changing the pigs’ susceptibility to Salmonella. How tissue colonization at slaughter may be being altered is yet unclear, as SNVs associated with both negative and positive isolation of Salmonella at slaughter were identified. Additionally, these variants are approximately 400 kbp from the DCDC2C gene, and thus the observed effect on the slaughter trait may be related, in part, to linkage with a functional SNV in a different gene. Regardless, DCDC2C may be a potential target for further investigation Salmonella susceptibility in pigs.

Analysis of the slaughter model identified an additional three intergenic variants near the gene encoding A-kinase anchor protein 12 (AKAP12) that were approaching the suggestive association threshold. An intron variant in AKAP12 was also identified in the top 15 significant SNVs for the slaughter trait. AKAP12 is a member of the AKAP family of scaffolding proteins involved in the recruitment and anchoring of protein kinases (PK) A and B. AKAP12 has been shown to regulate the subcellular localization of PKA and play a role in PKB regulation [62, 63]. PKA and PKB signaling play a crucial role in intramacrophage killing. Salmonella is known to subvert the host cell machinery and manipulate these pathways to persist intracellularly through the reduction of reactive oxygen intermediates or the manipulation of motor proteins [64, 65]. Studies have shown that inhibiting PKA and PKB activation greatly reduces virulence and impairs intracellular growth and survival of Salmonella [66, 67].

In light of the role that AKAP12 plays in the localization and regulation of PKA and PKB, which are key players in the intracellular survival of Salmonella, it is possible that variants near AKAP12 are altering the expression of AKAP12 and reducing or inhibiting the normal localization of PKA and PKB. This, in turn, may be impairing intracellular survival of Salmonella in host macrophages and preventing or limiting the systemic spread of Salmonella to lymphatic tissues and may account for the significantly higher frequency of these variants in the control pigs that were negative for Salmonella in tissues collected at slaughter.

GWAS analysis of seropositivity from nursery to finisher stage

GWAS analysis did not identify any significant associations with the tested variants. The closest SNV approaching the suggestive association threshold was an intergenic variant nearest to the gene RALGAPA2, which encodes the gene Ral GTPase-activating protein (RalGAP) catalytic alpha subunit 2 (α2) of RalGAP and was observed at a higher frequency in pigs that tested seropositive for Salmonella at least once. An additional two intergenic variants near and one intron variant in RALGAPA2 were identified in the top 15 significant SNVs for the seropositivity model. RalGAPs are involved in the negative regulation of RalGTPases, RaA and RaB, in the Ras/Ral signaling pathways and regulate crucial cellular processes including response to infection and mediation of inflammation [68–72]. Notably, constitutive expression of RalGTPases has been shown to promote tumorigenesis, increased expression of inflammatory cytokines, and increased epithelial permeability [69, 73–76].

It is possible that these variants are altering the expression of RALGAPA2 and, in turn, causing constitutive expression of RalGTPases. Considering the role of Ral GTPases in cell survival, inflammation, and permeability, this constitutive expression may be promoting susceptibility to Salmonella. Firstly, by creating an inflammatory environment in which Salmonella is known to thrive [77–79], or secondly, by impairing infected cell death and promoting the intracellular survival of Salmonella leading to chronic or persistent infection and higher antibody production [80–83]. However, it is important to note that the ELISA chosen for the current study tested for Salmonella IgM antibodies, but it is possible that some pigs had only produced a preliminary IgM response. Further studies are needed to assess the presence of IgM antibodies and whether they may result in stronger associations between the identified variants and Salmonella seropositivity. Additionally, a targeted study into RALGAPA2 and its pathways may identify an association with Salmonella susceptibility that was missed at the genome-wide level.

Conclusion

To the best of the authors knowledge, this is the first GWAS analysis assessing susceptibility to Salmonella in pigs at different stage of production on commercial farms and at slaughter. The variants identified herein were generally in or near genes involved in broad innate immunity and may be of great interest in improving genetic resistance to control enteric pathogens on swine farms. Overall, much of the porcine innate and humoral immune responses to these pathogens remains unknown. Further investigation of these genes on-farm, and the function, expression, and pathways of these genes both in vitro may greatly improve our understanding of the genetic basis of susceptibility to Salmonella and other swine enteric pathogens.

Methods

Animal use in this project was approved by the University of Guelph Animal Care Committee (AUP# 3124) and follows Canadian Council of Animal Care guidelines (CCAC, 2009).
Animals and sample collection
Farm and pig selection for this study have been detailed previously [84]. Briefly, 14 groups of 54–60 pigs (three-way Yorkshire x Landrace x terminal boar line cross) were selected from eight commercial farrowing sources in Southern Ontario; all pigs were housed on commercial farms and kept under the management of the farm personnel. Two groups of pigs were selected from each of six farrowing sources (designated Cohorts One and Two), while the remaining two farrowing sources had only one group each (Cohort One) for a total of 14 groups. During the nursery stage, pigs received either a conventional diet (high complexity) or a lower cost reduced animal protein diet (low complexity) in which animal protein was replaced by plant protein as part of a larger study [85]. After which, all pigs received conventional diets during grower-finisher stage.

Blood and fecal samples were collected at weaning, and at the end of the nursery, grower and finisher stages. Rectal swabs (Starplex®, VWR International, Mississauga, Ontario, Canada) were taken in the event that a fecal sample could not be obtained. Blood was drawn from the jugular or suborbital vein, transported to the lab in a cooled, and then centrifuged at 1500 × g for 20 min. Sera were stored at −20 °C. At the end of production pigs were shipped to the abattoir for slaughter where palatine tonsils and submandibular lymph nodes were collected from a subset of 580 pigs.

Salmonella antibody detection
Collected sera were tested for IgG antibodies to Salmonella O-antigens 1, 3, 4, 6, 7, 9, 10, and 12 via a commercial enzyme-linked immunosorbent assay (ELISA; pigtype® Salmonella Ab kit, QIAGEN Leipzig GmbH, Leipzig, Germany) as described previously [86]. Sample optical density (OD) was measured at 450 nm using a BioTek Synergy HT Multi-Mode Microplate Reader and BioTek’s Gen5 software version 11.1. The sample-to-positive (S/P) ratios were calculated as follows:

$$S/P = \frac{OD_{sample} - Mean\ OD_{negative\ control}}{Mean\ OD_{positive\ control} - Mean\ OD_{negative\ control}}$$

A sample was considered seropositive if the S/P ratio was greater or equal to 0.3.

Salmonella isolation
Fecal samples and tissues were cultured as detailed previously [84]. Briefly, 10 g of fecal or tissue sample was homogenized in 50 mL of tetrathionate broth (Oxoid, Nepean, Ontario, Canada) using a Seward Stomacher 400 Circulator (Seward Laboratory Systems Inc., Bohemia, New York, USA) and incubated for 24 h at 37 °C. After which, 0.1 mL of tetrathionate broth culture was transferred into 9.9 mL Rappaport Vassiliadis (RV) broth (Oxoid, Nepean, Ontario, Canada) followed by another 24 h incubation at 42 °C. Then, a loopful of RV culture was plated onto xylose-lysine-tergitol 4 agar (Becton Dickinson™, Baltimore, Maryland, USA) and incubated for a final 24 h at 37 °C. Salmonella colonies were confirmed by a Salmonella O Antiserum Poly A-I and Vi (Becton Dickinson™, Grayson, Georgia, USA) slide agglutination test.

Salmonella phenotypes
The following binary traits were defined for each pig: Salmonella seropositivity from end of nursery to end of finisher (never seropositive vs. seropositive once or more), Salmonella shedding from weaning to end of finisher (never shed vs. shed once or more), and positive isolation of Salmonella from tonsils and/or lymph nodes at slaughter (yes/no). Seropositivity at weaning was excluded when categorizing pigs into binary traits due to the concern of maternal antibodies confounding the results in the GWAS. Supplementary Figure 1 depicts a simplified flow chart of the sample collection and case-control groups by phenotype.

Genotyping and quality control
DNA was extracted from tail dockings, ear tissue, or blood using the DNEasy Blood and Tissue kit (Qiagen, Mississauga, Ontario Canada), with either 25 mg of tissue or 100 ul of blood [35]. DNA was genotyped at Eurofins BioDiagnostics, Inc. (River Falls, Wisconsin, USA) using a custom 54 K Affymetrix Axiom® myDesign™ chip designed in consultation with the Canadian Centre for Swine Improvement (Ottawa, Ontario, Canada). Quality control (QC) of SNVs was performed in PLINK v1.9 [87, 88]. As this chip contained select SNVs from multiple research groups, some of which had proprietary labels that precluded their identification in the genome, SNVs were removed if they had no corresponding rsID in Ensembl Sscrofa11.1. Pigs with SNV call rates of less than 90% were excluded and SNVs with a minor allele frequency lower than 5% or a call rate less than 95% were removed from further analysis. The SNV call rates were compared between cases and controls, and Fisher’s exact test was used to exclude any SNV in which the missingness between a case and control was significantly different ($p < 1.0 \times 10^{-5}$).

Genome-wide association study
A GWAS model was used to assess each of the phenotypes described above. Pigs that tested seropositive for Salmonella or shed Salmonella at least once on-farm, and pigs that were positive for isolation of Salmonella at slaughter were considered to be the case population. Pigs that never tested seropositive, shed Salmonella, or...
tested negative at slaughter were assigned as the control population for the GWAS models. The outcome variable for each model was seropositive at least once from end of nursery to end of finisher (seropositivity model), shedding detected at least once from weaning to end of finisher (shedding model), and positive isolation at slaughter (colonization model), respectively.

Univariable analysis of each phenotype was performed using a mixed-effects logistic regression model in Stata (Stata/IC 14.2 for Windows, StataCorp LP, Texas, USA) to assess the significance of associations between covariates and the phenotype ($p < 0.05$) and subsequent inclusion in the GWAS model [84, 86]. For the seropositivity model, only farm was included as a covariate. Covariates included in the shedding model were diet and cohort, while covariates for the colonization model were age at slaughter, cohort, and season. To account for the cryptic relatedness of the pigs in a study population with no sire information, a genomic relatedness matrix (GRM) was calculated using the genome-wide efficient mixed model association algorithm (GEMMA) v0.96 [37, 38, 89]. Genome-wide analyses of case-control data were performed using a generalized logistic mixed model association test (GMMAT) [90] using the GMMAT v0.9.3 package for R [91, 92]. The covariates analyzed in Stata and the GRM created in GEMMA were included in the GMMAT analysis. The chosen $p$-value threshold for suggestive and significant associations was $1.0 \times 10^{-5}$ and $5.0 \times 10^{-7}$, respectively based on a study by Burton et al. [39]. The Wald test was used to assess the significance of associations between SNVs and the Salmonella phenotypes. Allele frequencies in case-control populations were calculated and an allelic chi-squared ($\chi^2$) test in PLINK v1.9 was used to determine if there was a significant difference between case and control pig populations. In the event that an intergenic variant was identified as significant, the closest gene within a 1 Mbp window, either upstream or downstream, was assumed to be the causative gene. The data generated from this GWAS has been deposited in the Animal Quantitative Trait Loci Database (AnimalQTLdb) and can be accessed as a group via [https://www.animalgenome.org/QTldb_supp/?t=MwAn101Xia](https://www.animalgenome.org/QTldb_supp/?t=MwAn101Xia), or individually via association (QTL) numbers: 194531–194535.

**Abbreviations**

AKAP: A-kinase anchoring protein; ELISA: Enzyme-linked immunosorbent assay; DCDC2C: Doublecortin domain containing 2C; GWAS: Genome-wide association study; GRM: Genomic relatedness matrix; GMMAT: Generalized logistic mixed model association test; GEMMA: Genome-wide efficient mixed model association; IgG: Immunoglobulin G; MBL: Mannan-binding lectin; NOD: Nucleotide-binding oligomerization domain-containing protein; OD: Optical density; PK: Protein kinase; PTP: Protein tyrosine phosphatase; PTPRU: Protein tyrosine phosphatase receptor type J; QC: Quality control; RalGDP: Ral GTPase-activating protein; RALGAP: Ral GTPase-activating protein catalytic alpha subunit 2; RV: Rappaport Vassiliadis; S.: Salmonella; SNV: Single-nucleotide variant; S/P: Sample-to-positive; ST6GALNAC3: ST6 (α-N-acetylgalactosaminyl-2,3-β-galactosyl-H-3)-N-acetylgalactosamine-α-2,6-sialyltransferase 3; TLR: Toll-like receptor

**Acknowledgements**

The authors would like to thank Saraany Nair, Jane Newman, and Dr. Jutta Hammermueller for their assistance in sample collection and processing, as well as Synergy Swine Inc., Conestoga Meat Packers, and the pork producers that participated in this study.

**Authors’ contributions**

The project was conceived, study designed, and funding obtained by BNL, AF, and RMF. Bacterial culture and host DNA extraction was performed by MHAG. Antibody assays were performed by CHS. Bioinformatics and statistical analysis was performed by CHS, RSF, and AF. CHS, AF, and BNL were major contributors to writing the manuscript. All authors read and approved the final version of the manuscript.

**Funding**

This work was supported by the Natural Sciences and Engineering Research Council of Canada Collaborative Research and Development Grant [grant number 4988429–15], the Ontario Ministry of Agriculture, Food and Rural Affairs [grant number OuroG2013–1472, HQP Fellowship for C. Schut]; Swine Innovation Porc [1239e], the Canadian Center for Swine Improvement, Alliance Genetics Canada, and Ontario Pork [15/015]. Alliance Genetics provided assistance in farm recruitment and sample collection. Genotyping was done on a custom SNP chip designed by the Canadian Centre for Swine Improvement, Inc. Funding sources did not contribute in any other way to the manuscript.

**Availability of data and materials**

The data generated from this GWAS has been deposited in the Animal Quantitative Trait Loci Database (AnimalQTLdb) and can be accessed as a group via [https://www.animalgenome.org/QTldb_supp/?t=MwAn101Xia](https://www.animalgenome.org/QTldb_supp/?t=MwAn101Xia), or individually via association (QTL) numbers: 194531–194535.

**Ethics approval and consent to participate**

Animal use in this project was approved by the University of Guelph Animal Care Committee (AUP# 3124) and follows Canadian Council of Animal Care guidelines (CCAC, 2009). Informed verbal consent was obtained from participating producers by phone and on-farm during a questionnaire of farm operating procedures as approved by the ethics committee.

**Consent for publication**

Not applicable.

**Competing interests**

None of the authors have any competing interests.

**Author details**

1Department of Pathobiology, University of Guelph, 50 Stone Rd E, Guelph, ON N1G 2W1, Canada. 2Department of Population Medicine, University of Guelph, Guelph, Ontario, Canada. 3Present address: Department of Pathology and Microbiology, Atlantic Veterinary College, University of PEI, Charlottetown, Prince Edward Island, Canada.

**Supplementary information**

Supplementary information accompanies this paper at [https://doi.org/10.1186/s12917-020-02344-9](https://doi.org/10.1186/s12917-020-02344-9).

Additional file 1: Figure S1. Visual representation of the sampling and processing of sera and fecal/tissue samples from pigs on-farm and at slaughter and the case-control population breakdowns of the three binary Salmonella phenotypes used in GWAS.

Additional file 2: Table S1. Allele frequencies in case and control pig populations for SNVs with significant associations in GWAS models.
64. Kashina AS, Semenova IV, Ivanov PA, Potekhina ES, Zalapin I, Rodionov VT. Protein kinase A, which regulates intracellular transport, forms complexes with molecular motors on organelles. Curr Biol. 2004;14(20):1877–81 [cited 2018 Nov 3] Available from: https://www.sciencedirect.com/science/article/pii/S0960982204007936?via%3Dihub.

65. Gallois A, Klein JR, Allen LA, Jones BD, Nauseef WM. Salmonella pathogenicity island 2-encoded type III secretion system mediates exclusion of NADPH oxidase assembly from the phagosomal membrane. J Immunol. 2001;166(9):5741–8 [cited 2018 Nov 3] Available from: http://www.ncbi.nlm.nih.gov/published/11313417.

66. Kuijl C, Savage ND, Marsman M, Tuin AW, Jansen L, Engan DA, et al. Intracellular bacterial growth is controlled by a kinase network around PKB/ Akt1. Nature. 2007;450(7170):75–32 [cited 2018 Nov 3] Available from: http://www.nature.com/articles/nature06345.

67. Uchhya K, Niki T. Salmonella enteritidis serovar Typhimurium infection induces cyclooxygenase 2 expression in macrophages: involvement of Salmonella pathogenicity island 2. Infect Immun. 2004;72(12):6800–9 [cited 2018 Nov 3] Available from: http://www.ncbi.nlm.nih.gov/published/15557607.

68. Moghadam AR, Patrad E, Tafseri E, Peng W, Fangman B, Pluard TJ, et al. Ral signaling pathway in health and cancer. Cancer Med. 2017;6(2):2998–3013 [cited 2018 Nov 1] Available from: http://doi.wiley.com/10.1002/cam4.1105.

69. Bodemann BO, Orvedahl A, Cheng T, Ram RR, Ou Y-H, Formstecher E, et al. Ras history. Small GTPases. 2010;1(1):2 [cited 2018 Oct 29] Available from: http://www.ncbi.nlm.nih.gov/published/21243894.

70. Cox AD, Der CJ. Ras history. Small GTPases. 2001;276(48):44504 [cited 2018 Oct 29] Available from: http://doi.wiley.com/10.4161/sgtp.1.1.12178.

71. Shirakawa R, Fukai S, Kawato M, Higashi T, Kondo H, Ikeda T, et al. Tuberous sclerosis tumor suppressor complex-like complexes act as GTPase-activating effectors SseF and SseG inhibit Rab1A-mediated autophagy to facilitate intracellular bacterial survival and replication. J Biol Chem. 2018;293(25):9662–73 [cited 2018 Nov 1] Available from: http://www.ncbi.nlm.nih.gov/published/29101724.

72. Feig LA. Ral-GTPases: approaching their 15 minutes of fame. Trends Cell Biol. 2003;13(8):383–91 [cited 2018 Nov 1] Available from: http://www.ncbi.nlm.nih.gov/published/12754967.

73. Wang XQ, Sun P, Paller AS. Inhibition of integrin-linked kinase/protein kinase B/Akt signaling: mechanism for ganglioside-induced apoptosis. J Biol Chem. 2001;276(48):44504–11 [cited 2018 Nov 3] Available from: http://www.ncbi.nlm.nih.gov/published/11570896.

74. Peschard P, McCarthy A, Leblanc-Dominguez V, Yeo M, Guichard S, Stamp 25 [cited 2018 Nov 1] Available from: https://www.ncbi.nlm.nih.gov/published/30863385.

75. 81. Ainslie-Garcia MH, Farzan A, Newman JE, Friendship RM, Lillie BN. Salmonella fecal shedding in pigs from birth to market and its association with the presence of Salmonella in palatine tonsils and submandibular lymph nodes at slaughter. Can J Vet Res. 2018;82(4):249–55 [cited 2018 Nov 4] Available from: http://www.ncbi.nlm.nih.gov/published/30863385.

76. Reinhardt H, Shoveller AK, Farzan A, McBride B, de Lange CFM, Huber L. Effect of nursery feeding program on serum haptoglobin, growth performance, and carcass characteristics of pigs reared on commercial farms. Can J Vet Res. 2018;83(4):255–60.

77. Schut CH, Farzan A, Ainslie-Garcia MH, Friendship RM, Lillie BN. Antibody responses to Salmonella in pigs from weaning up to marketing and presence of Salmonella at slaughter. Foodborne Pathog Dis. 2019;16(13):985–94 [cited 2019 Mar 28] Available from: https://www.liebertpub.com/doi/10.1089/fpd.2018.2454.

78. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK rising to the challenge of larger and richer datasets. Gigascience. 2015;4(1):7 [cited 2018 Oct 29] Available from: http://www.ncbi.nlm.nih.gov/published/25722852.

79. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81(3):559–75 [cited 2018 Oct 29] Available from: http://linkinghub.elsevier.com/retrieve/pii/S0002929207008328.

80. Weir BS, Anderson AD, Hepler AB. Genetic relatedness analysis: modern data and new challenges. Nat Rev Genet. 2006;7(10):771–80 [cited 2018 Oct 29] Available from: http://www.ncbi.nlm.nih.gov/published/12809640.

81. Chen H, Wang C, Conomos PM, Stilp AM, Li Z, Sofer T, et al. Control for population structure and relatedness for binary traits in genetic association studies via logistic mixed models. Am J Hum Genet. 2016;98(4):653–66 [cited 2018 Sep 24] Available from: http://www.ncbi.nlm.nih.gov/published/27018471.

82. Chen H, Conomos MP. GMMA: Generalized linear mixed model association tests. R package version 0.9.3. 2018 [cited 2018 Nov 10] Available from: https://rdrr.io/github/hanchenphd/GMMA/

83. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2018. [cited 2018 Nov 10] Available from: https://www.R-project.org/.

Publisher’s Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.