





# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
<b>2</b>	<b>Grundlagen genomische Selektion</b>	<b>1</b>
2.1	Was versteht man unter genomischer Selektion? . . . . .	1
2.2	SNP-Marker . . . . .	1
2.3	Vom Marker zum Gen . . . . .	2
2.4	Vom Marker zum genomischen Zuchtwert . . . . .	3
2.5	Sicherheit von genomischen Zuchtwerten . . . . .	4
2.6	Genomisch optimierte Zuchtwerte . . . . .	5
<b>3</b>	<b>Was können genomische Zuchtwerte?</b>	<b>7</b>
3.1	Genomische Zuchtwerte sind genauer als Abstammungszuchtwerte . . . . .	7
3.2	Genomische Zuchtwerte können sich ändern . . . . .	8
3.3	Genomische Zuchtwerte können Vollgeschwister ohne Leistungs- informationen unterscheiden . . . . .	9
3.4	Genomische Zuchtwerte können genetisch wertvolle Tiere und Linien identifizieren . . . . .	9
<b>4</b>	<b>Logistik bei der genomischen Selektion</b>	<b>10</b>
4.1	Was muss man tun? . . . . .	10
4.2	Was bekommt man? . . . . .	12
4.3	Was geschieht im Hintergrund? . . . . .	12
<b>5</b>	<b>Imputing</b>	<b>13</b>
5.1	Einleitung . . . . .	13
5.2	Wie funktioniert Imputing? . . . . .	13
5.3	Genauigkeit des Imputings . . . . .	16
5.4	Genomische Zuchtwertschätzung mit imputierten Genotypen	17
<b>6</b>	<b>Aktuelle Forschungsarbeiten</b>	<b>19</b>
6.1	Imputing vom 50k-Chip auf den BovineHD-Chip . . . . .	19
6.2	Genomische Zuchtwerte für kleine Rassen . . . . .	20

6.3	Nutzung von Sequenzdaten in der genomischen Zuchtwertschätzung . . . . .	21
6.4	Nutzung von typisierten Kühen in der genomischen Selektion	23

# 1 Einleitung

Das Ziel der Tierzucht besteht darin, die genetische Veranlagung von Tieren zu schätzen und im Hinblick auf die entsprechende Nutzung zu verändern bzw. zu verbessern. Bis vor wenigen Jahren konnte das genetische Potenzial, bzw. der Zuchtwert eines Tieres nur über seine eigenen Leistungen und/oder über die Leistungen von verwandten Tieren bestimmt werden. Ideal wäre es jedoch, wenn die genetische Veranlagung der Tiere direkt aus den Erbinformationen abgelesen werden könnte. Die genomische Selektion kommt dieser Wunschvorstellung sehr nahe.

In dieser Broschüre soll die genomische Selektion von den theoretischen Grundlagen, über die praktische Anwendung bis hin zu zukünftigen Entwicklungen genauer erläutert werden.

## 2 Grundlagen genomische Selektion

### 2.1 Was versteht man unter genomischer Selektion?

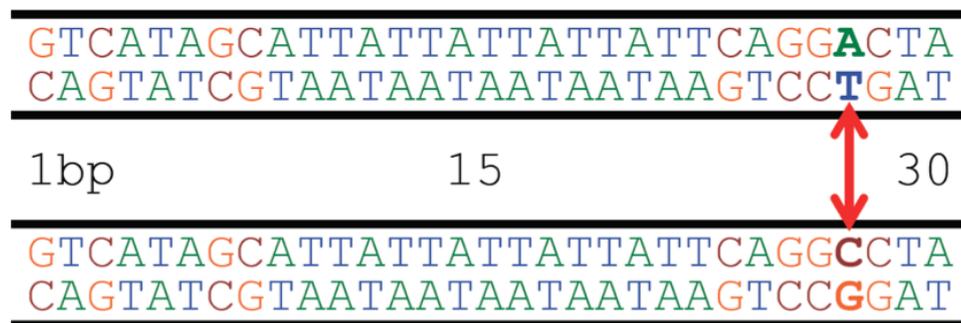
Unter dem Begriff genomische Selektion werden Ansätze zusammengefasst, die das Ziel haben, Informationen aus dem Erbgut direkt für die Zuchtwertschätzung und weitere Anwendungen (u.a. Erbfehlertest) zu nutzen. Diese Möglichkeiten sind in erster Linie Entwicklungen in der Molekulargenetik zu verdanken. Mit immer effizienteren Verfahren wurde es möglich, Informationen aus dem Erbgut in grossem Umfang und höherer Auflösung vergleichsweise kostengünstig zur Verfügung zu stellen.

### 2.2 SNP-Marker

Obwohl das Genom des Rindes inzwischen weitgehend entschlüsselt ist, ist für die Mehrzahl der züchterisch interessanten Merkmale wenig über die Lage und Wirkung der Gene bekannt. Aus diesem Grund arbeitet die genomische Selektion nicht direkt auf der Basis von Genen, sondern auf der Basis von Markern. Der Grundbaustein der genomischen Selektion sind sogenannte SNP (single nucleotide polymorphism)-Marker, welche mehr oder weniger

gleichmässig über das gesamte Erbgut eines Tieres verteilt sind. Der grosse Vorteil dieser Marker ist, dass sie in sehr hoher Anzahl vorkommen. Man rechnet heute damit, dass im Genom von Rindern etwa 25-30 Millionen SNP existieren.

SNP sind punktuelle Veränderungen an bestimmten Positionen im Genom (Abbildung 1). Ein SNP ist nur einen genetischen Buchstaben (bp) lang und es gibt jeweils nur zwei unterschiedliche Varianten in der Population. Die Ausprägung eines SNP wird im Rahmen einer so genannten Typisie-



**Abbildung 1:** SNP als punktuelle Veränderung im Genom

rung festgestellt. Beim Typisieren werden einzig die genetischen Zustände des untersuchten Tieres an den SNP gelesen. Werden viele SNP gleichzeitig analysiert, verwendet man dafür in der Regel Chips.

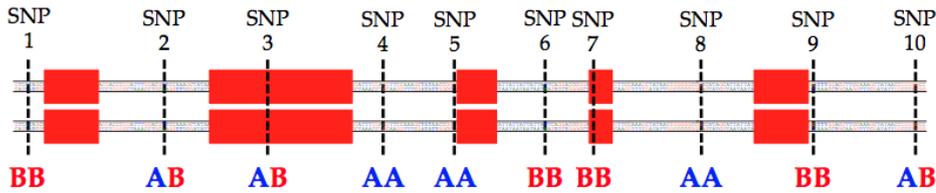
In der Schweiz werden Chips der Firma Illumina verwendet. Mit den beiden derzeit in der Routine verwendeten Chip-Typen können rund 54'000 SNP (BovineSNP50 v2 = Bovine50k-Chip) oder rund 19'000 SNP (Gene-Seek Genomic Profiler for Dairy-LD = LD-Chip) typisiert werden. Daneben existieren weitere Varianten, die je nach Anwendung eingesetzt werden können. Selbst die Charakterisierung des gesamten Genoms im Rahmen einer so genannten Sequenzierung aller 3 Milliarden Basen ist heute zu vernünftigen Preisen möglich.

## 2.3 Vom Marker zum Gen

Durch die gleichmässige Verteilung von sehr vielen SNP über das gesamte Genom geht man davon aus, dass SNP in der Nähe von Genen liegen, die

züchterisch relevante Merkmale beeinflussen (Abbildung 2).

Basierend auf der statistisch-genetischen Grösse des Kopplungsungleichgewichts (linkage disequilibrium LD) versucht man in der genomischen Zuchtwertschätzung, die Geneffekte indirekt über die SNP-Effekte zu schätzen. Je vollständiger das LD ist, umso genauer werden über die Analyse der Marker die kausalen (ursächlichen) Geneffekte erfasst.



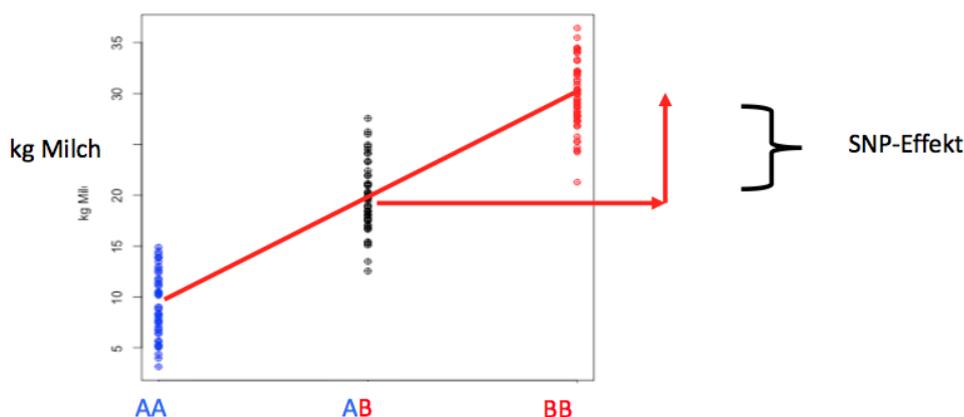
**Abbildung 2:** SNP-Marker (AA, AB, BB) liegen in der Nähe von Genen (rote Zylinder)

## 2.4 Vom Marker zum genomischen Zuchtwert

Um herauszufinden, welche SNP-Marker wie viel Einfluss auf ein Merkmal haben, muss ein Zusammenhang zwischen den Typisierungsergebnissen (Genotypen) und den Leistungen, bzw. konventionellen Zuchtwerten, hergestellt werden. Dies geschieht in der Effektschätzung, für die in der Regel typisierte Stiere mit einem sicheren Nachzuchtprüfungsergebnis (Trainings-Tiere) verwendet werden. Je mehr sicher geprüfte Stiere zur Verfügung stehen, desto besser kann man die Zuordnung eines SNP zu einem Merkmal vornehmen und die Höhe des Einflusses bestimmen (Effektschätzung).

Die Abbildung 3 zeigt schematisch den Zusammenhang zwischen der Ausprägung eines bestimmten SNP-Markers und dem Merkmal Milch kg. Daraus ist ersichtlich, dass Stiere mit B im Durchschnitt besser sind als Stiere mit A. Der Effekt von einem B beträgt rund +10 kg.

In der Abbildung 4 wird der Verlauf der Effektschätzung und die Berechnung der direkt genomischen Zuchtwerte (DGZW) schematisch dargestellt: Mittels einer Gegenüberstellung von konventionellen Zuchtwerten (CH oder Interbull) und Typisierungsdaten von nachzuchtgeprüften Stieren werden für sämtliche typisierten SNP-Marker ein Effekt geschätzt. Für die Berechnung



**Abbildung 3:** Effekt der SNP-Variante B auf die Milchleistung (als Beispiel)

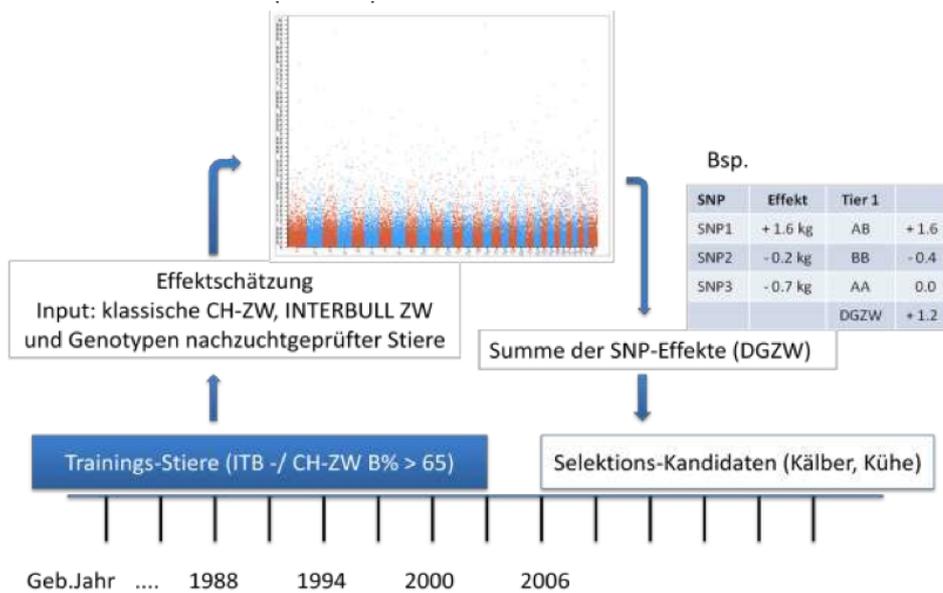
der DGZW werden diese Effekte für ein bestimmtes Tier über alle SNP-Marker aufsummiert. Dies ist für alle beliebigen (jüngeren) Tiere, die selber noch keine eigenen konventionellen Zuchtwertinformationen aufweisen, möglich.

## 2.5 Sicherheit von genomischen Zuchtwerten

Die Sicherheit der genomischen Zuchtwerte hängt in erster Linie von der Grösse und der Struktur des Trainingsdatensatzes ab. Eine optimale Struktur liegt dann vor, wenn die Genetik von Tieren, deren DGZW geschätzt werden soll, über möglichst viele verwandte Tiere im Trainingsdatensatz vertreten ist.

Die wichtigsten nachzuchtgeprüften Stiere, die für den Trainingsdatensatz in Frage kommen, sind heute typisiert. Bei der Rasse Holstein ist der Zugang zu den Typisierungen vieler Stiere allerdings nur beschränkt möglich, da der internationale Austausch dieser Daten schwierig ist. Bei der Rasse Braunvieh können die Genotypen innerhalb der Inter-genomics-Länder für Routineauswertungen von allen Beteiligten genutzt werden. Neben dem internationalen Austausch der Stieren-Genotypen bildet die Integration der Kuh-Genotypen in den Trainingsdatensatz eine weitere Möglichkeit, diesen zu vergrössern.

Die Sicherheit genomischer Zuchtwerte leitet man aus der Übereinstim-



**Abbildung 4:** Ablauf der Effektschätzung und Berechnung der direkt genomischen Zuchtwerte (DGZW)

mung zwischen DGZW und Zuchtwerten aus der Nachzuchtprüfung ab. Dafür wird ein Teil der Stiere im ursprünglichen Trainingsdatensatz von der Effektschätzung ausgeschlossen (Validierungs-Tiere). Mit den restlichen Stieren wird dann eine separate Effektschätzung durchgeführt. Die Sicherheit der DGZW aller Tiere resultiert aus der Übereinstimmung zwischen dem DGZW und den Zuchtwerten aus der Nachzuchtprüfung der Validierungs-Tiere.

In der Tabelle 1 sind für die drei Schweizer Milchviehhaupttrassen die Sicherheiten der DGZW zusammengestellt.

## 2.6 Genomisch optimierte Zuchtwerte

Der DGZW eines Tieres stellt eine zusätzliche Information zu den bisherigen traditionellen Zuchtwerten dar. Die bestmögliche Schätzung für den genetischen Wert eines Tieres ergibt sich, wenn alle vorliegenden Informationen (traditionell und genomisch) genutzt und zu einem Index zusammengefasst werden.

Diesen kombinierten Wert nennt man genomisch optimierter Zuchtwert (GOZW).

**Tabelle 1:** Sicherheit (B%) der DGZW für ausgewählte Merkmale

<b>Merkmal</b>	<b>RH, SF</b>	<b>BV</b>	<b>HO</b>
<i>Produktion</i>	56-59%	55%	60%
<i>Zellzahl</i>	34%	55%	43%
<i>Exterieur</i>	8-65%	32-75%	22-61%
	ø33.8%	ø56.0%	ø40.7%
<i>Fruchtbarkeit</i>	56-61%	50-58%	32-42%
<i>Nutzungsdauer</i>	26%	30%	30%
<i>Geburtsablauf</i>	17-27%		27-29%

Je nach Alter des Tieres bildet der DGZW zusammen mit einem Abstammungszuchtwert bei Jungtieren, bzw. dem traditionellen Zuchtwert bei Tieren mit Leistungsinformationen, den GOZW des Tieres. Die Gewichtung von genomischem und traditionellem Zuchtwert im GOZW erfolgt anhand der Sicherheiten der beiden Werte. Bei jungen Tieren mit einem (unsicheren) Abstammungszuchtwert erhält der DGZW ein relativ grosses Gewicht. Sobald ein Tier aufgrund von Eigen- oder Nachkommenleistungen einen traditionellen Zuchtwert mit einer höheren Sicherheit aufweist, wird das Gewicht des DGZW im GOZW kleiner. Daher entspricht der GOZW bei sicher nachzuchtgeprüften Stieren beinahe dem traditionellen Zuchtwert. Die Nomenklatur des GOZW richtet sich danach, welcher traditionelle Zuchtwert mit dem DGZW kombiniert wurde:

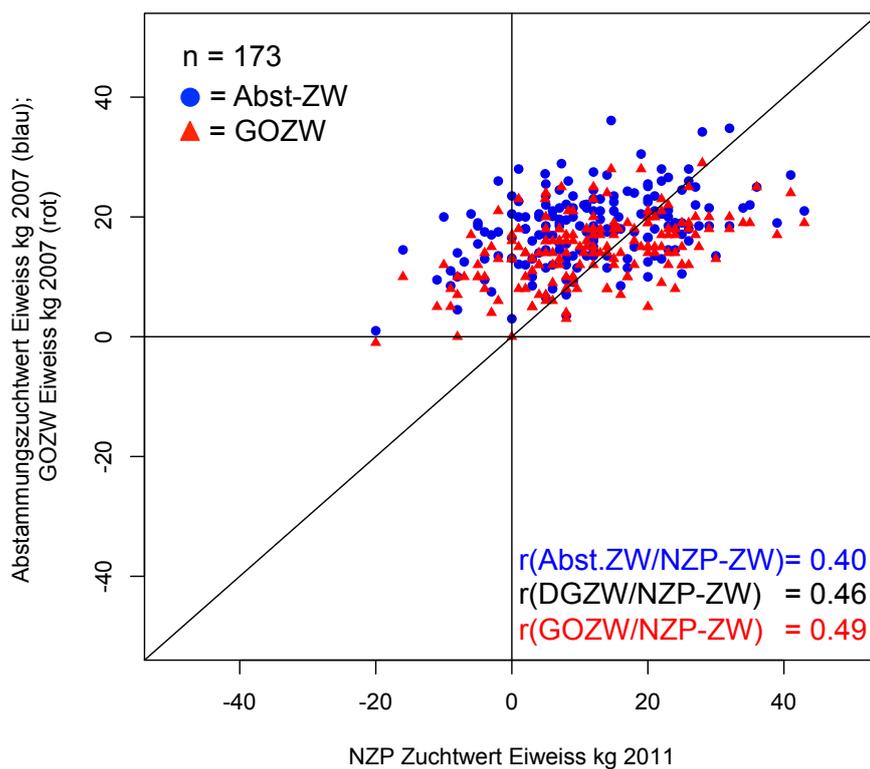
#### **Deklaration der GOZW**

DGZW	+ Abstammungszuchtwert	=	<b>GA</b>
DGZW	+ CH-Nachzuchtprüfungsresultat	=	<b>G</b>
DGZW	+ Interbull-Zuchtwert	=	<b>GI</b>

### 3 Was können genomische Zuchtwerte?

#### 3.1 Genomische Zuchtwerte sind genauer als Abstammungszuchtwerte

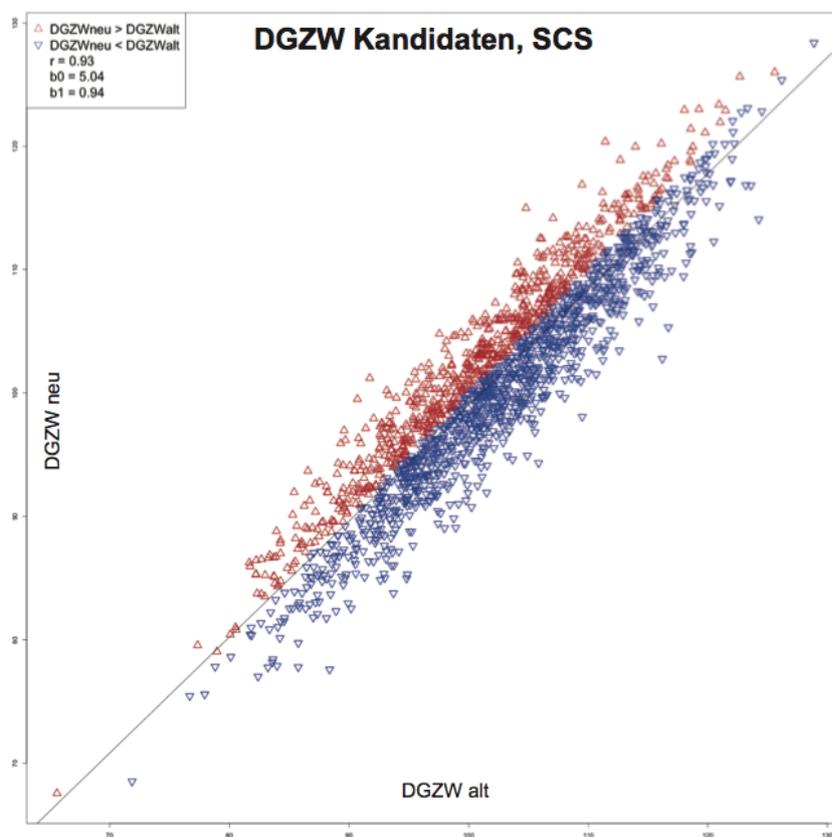
Genomische Zuchtwerte stimmen mit realisierten Zuchtwerten aus der Nachzuchtprüfung (NZP) besser überein als Abstammungszuchtwerte. Dies konnte im Rahmen der Validierung für Interbull gezeigt werden. Die Abbildung 5 zeigt die Beziehung zwischen dem aktuellen Nachzuchtprüfungsergebnis und dem Abstammungszuchtwert bzw. GOZW aus dem Jahr 2007. Selektiert man Tiere an Hand genomischer Zuchtwerte, trifft man eine genauere Entscheidung und erzielt einen grösseren Züchterfolg.



**Abbildung 5:** Validierung nach Interbull-Kriterien am Beispiel der Rasse Braunvieh

## 3.2 Genomische Zuchtwerte können sich ändern

Der DGZW eines Tieres basiert auf rund 50'000 Genotypen. Diese Information aus dem Erbgut ist unveränderlich. Dagegen aber sind die Effekte aus der Effektschätzung abhängig vom Trainingsdatensatz. Jeder Lauf der herkömmlichen Zuchtwertschätzung liefert zusätzliche Informationen für die Effektschätzung. Daraus ergeben sich Änderungen bei den Effekten, die in leicht veränderten DGZW sichtbar werden (Abbildung 6).



**Abbildung 6:** DGZW von Kandidaten (Stiere ausserhalb Trainingsdatensatz) aus Effektschätzungen 04/2012(alt) und 08/2012(neu)

### **3.3 Genomische Zuchtwerte können Vollgeschwister ohne Leistungsinformationen unterscheiden**

DGZW basieren ausschliesslich auf Informationen aus dem Erbgut. Bei der Paarung verschmelzen 50% väterliches und 50% mütterliches Erbgut miteinander und bilden das Genom des Nachkommen. Diese Mischung hat zur Folge, dass Vollgeschwister genetisch verschieden sind. Unterschiede im Erbgut führen zu unterschiedlichen DGZW. Vollgeschwister sind daher in einer genomischen Zuchtwertschätzung unterscheidbar, ohne dass sie Leistungen erbringen müssen.

### **3.4 Genomische Zuchtwerte können genetisch wertvolle Tiere und Linien identifizieren**

Die genomische Selektion eröffnet die Möglichkeit, aus ungewöhnlichen Abstammungen oder bisher wenig in Erscheinung getretenen Kuhfamilien Tiere mit überdurchschnittlicher genetischer Veranlagung zu finden. Auch in Betrieben, die bis anhin züchterisch kaum aktiv waren, können dank den genomischen Zuchtwerten Kühe identifiziert werden, die für die Gesamtpopulation von züchterischem Nutzen sein können. Des Weiteren kann sich auch die Genotypisierung von weiblichen Jungtieren lohnen: Die genomischen Zuchtwerte sind genauer als die Abstammungszuchtwerte. Will der Züchter den sichersten Weg einschlagen, ist eine Typisierung angezeigt. Diese kann in den ersten Lebenswochen eines Kuhkalbes gemacht werden, um über Aufzucht oder Mast zu entscheiden. Oder das Jungrind kann vor der ersten Besamung typisiert werden, um den richtigen Paarungspartner zu finden. Durch die Typisierung werden Zuchtwerte von zusätzlichen Merkmalen im Bereich der Fitness und des Exterieurs verfügbar, die bei der Wahl des passenden Besamungstieres hilfreich sein können. Genomische Zuchtwerte können zudem eine Entscheidungsgrundlage für den Einsatz von gesextem Spermia sein oder ob das Jungtier allenfalls für Embryotransfer (ET) gespült werden soll.

## 4 Logistik bei der genomischen Selektion

### 4.1 Was muss man tun?

Falls ein Züchter eines seiner Tiere genomisch untersuchen lassen möchte, ist der erste Schritt die Entnahme von geeignetem Probenmaterial. In der Schweiz sind dabei Haare sowie bei KB-Stieren Samendosen üblich. Die Entnahme der Probe muss durch eine betriebsfremde Person (Milchkontrolleur, Besamungstechniker, Tierarzt usw.) erfolgen. Der Probenentnahme sollte besondere Beachtung geschenkt werden, da nur eine genügende Menge an Zellmaterial zu qualitativ guten Typisierungsergebnissen führt. Folgende Punkte sind bei der Entnahme von Haarproben wichtig:

- Haare
  1. Am besten eignen sich die festen dicken Haare aus der Schwanzquaste (siehe Abbildung 7).
  2. Das Vermischen mit fremden Tierhaaren muss verhindert werden.
  3. Die Haare müssen sauber und trocken sein.
  4. Die Haare müssen gezupft und dürfen nicht abgeschnitten werden. Nur die Haarwurzeln enthalten die DNA des Tieres.
  5. Die Haarproben werden am besten in ein Papiercouvert verpackt. Papiercouverts eignen sich besser als Plastikbeutel, da das Risiko, dass die Probe feucht wird, kleiner ist.
- Bei Samendosen muss beachtet werden, dass die Dose beim Versand nicht knickt. Dies kann beispielsweise dadurch verhindert werden, dass die Dose an einen dickeren Karton geklebt wird.

Der zweite Schritt auf dem Weg zu genomischen Zuchtwerten beinhaltet die Beschaffung und das Ausfüllen eines Auftragsformulars. Diese Formulare können auf den Webseiten der Milchviehzuchtverbände abgerufen werden:

*Braunvieh Schweiz:*

<http://homepage.braunvieh.ch/documents/Auftragsformular-Genomische-Selektion-D.pdf>



**Abbildung 7:** Haare am besten aus der Schwanzquaste ausreissen

*Schweizerischer Holsteinzuchtverband:*

[http://www.holstein.ch/medias/pdf/genetique/genomique/bestellungsformular\\_gs.pdf](http://www.holstein.ch/medias/pdf/genetique/genomique/bestellungsformular_gs.pdf)

*Swissherdbook:*

[http://www.swissherdbook.ch/fileadmin/customer/02\\_Genetik/22-Zuchtwerte/222-Genom-ZW/Auftragsformular.pdf](http://www.swissherdbook.ch/fileadmin/customer/02_Genetik/22-Zuchtwerte/222-Genom-ZW/Auftragsformular.pdf)

Der Auftraggeber hat für die Typisierung die Möglichkeit, aus mehreren Chip-Typen zu wählen. Dies muss auf dem Auftragsformular angegeben werden. Aktuell bestehen folgende Möglichkeiten:

- Low Density Chip
- Medium Density Chip (50k)

Aus der Sicht des Züchters sind diese beiden Chips in etwa gleichwertig. Der einzige Unterschied besteht darin, dass GOZW, die auf dem Low Density Chip (LD-Chip) basieren, etwas geringere Sicherheiten aufweisen als GOZW, die auf dem Medium Density Chip basieren. Neben der Untersuchung von SNP bietet der LD-Chip die Möglichkeit, genetische Zusatztests und Erbfehler tests durchzuführen. Beispiele für solche Tests sind Rotfaktor, Hornlosigkeit, Kappa-Kasein, CVM, Brachyspina, SMA und Weaver. Für einige dieser

Zusatztests ist eine Lizenzgebühr zu entrichten. Aus diesem Grund muss auf dem Auftragsformular angegeben werden, welche Zusatztests durchgeführt werden sollen. Es ist zu beachten, dass gewisse Tests nicht bei allen Rassen Sinn machen. Das ausgefüllte Auftragsformular muss zusammen mit der Probe an Qualitas AG in Zug gesendet werden. Wichtige Informationen zur Probenentnahme sowie Adressen und die aktuellen Preise sind auf den Auftragsformularen zu finden.

## 4.2 Was bekommt man?

Rund 40 bis 70 Tage nach dem Einsenden der Probe erhalten der Auftraggeber und der Tierbesitzer vom zuständigen Zuchtverband die Resultate der genomischen Zuchtwertschätzung per Post. Diese Resultate beinhalten DGZW und GOZW für Produktions- und Exterieur-Merkmale sowie für Persistenz und Zellzahl. Je nach Zuchtverband werden auch noch die Zuchtwerte für zusätzliche Fitnessmerkmale (Nutzungsdauer, Melkbarkeit, Fruchtbarkeit usw.) veröffentlicht. Dreimal jährlich, jeweils nach einer konventionellen Zuchtwertschätzung, werden die GOZW aller typisierten Tiere neu berechnet und auf den Datenbanken der Zuchtverbände gespeichert. Ab dann erscheinen die GOZW auch auf den Abstammungsausweisen und sind somit offiziell publiziert.

## 4.3 Was geschieht im Hintergrund?

Die bei Qualitas AG eingetroffenen Proben werden gekühlt gelagert und an jedem ersten Dienstag im Monat gebündelt an die Laborfirma GeneSeek® in den USA geschickt. Dort wird zuerst die DNA aus den Proben extrahiert und anschliessend wird diese DNA mit einer Apparatur der Firma Illumina an der gewünschten Anzahl SNP untersucht. Dieser Vorgang wird als SNP-Genotypisierung bezeichnet. Die dabei entstehenden Daten werden zurück an Qualitas AG gesandt. Bei Qualitas AG werden die SNP-Genotypen aufbereitet und falls notwendig imputiert (vgl. Kapitel 5). Danach werden sie auf den Datenbanken der Zuchtverbände gespeichert, wo anschliessend die Schätzungen der DGZW und GOZW durchgeführt werden. Die für die Schätzung der

DGZW nötigen SNP-Effekte werden nach jeder Routine-Zuchtwertschätzung neu berechnet.

## 5 Imputing

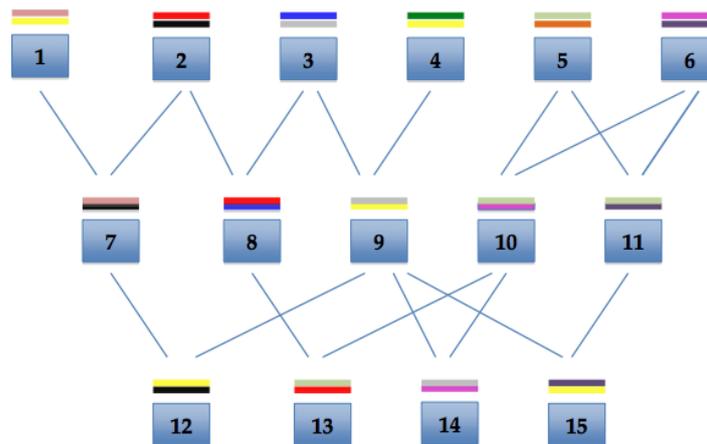
### 5.1 Einleitung

In der Schweiz stehen der Bovine50k-Chip und der LD-Chip für die Typisierung zur Verfügung. Der Bovine50k-Chip enthält etwa 54'000 SNP, der LD-Chip etwa 19'000. Der Bovine50k-Chip ist für die Effektschätzung und die Berechnung der DGZW Standard. Damit es durch die Verwendung des LD-Chips nur zu kleinen Qualitätseinbussen bei den DGZW kommt, müssen die Typisierungsergebnisse vom LD-Chip mit einem speziellen Verfahren erweitert werden. Dieses Verfahren wird als Imputing bezeichnet. Im Allgemeinen werden unter dem Begriff Imputing Verfahren zusammengefasst, mit denen der Genotyp eines grossen Chip (z.B. Bovine50k) ausgehend vom Genotyp eines kleinen Chip (z.B. LD) vorausgesagt werden kann. Imputing funktioniert sogar auch bei nicht typisierten Tieren, falls diese mit anderen typisierten Tieren eng verwandt sind.

Durch eine verstärkte Verwendung von LD-Chips kann die wirtschaftliche Effizienz der genomischen Selektion gesteigert werden, da bei gleichem finanziellen Mitteleinsatz mehr Kälber typisiert werden können. Zudem wird damit die genomische Selektion auch bei Kühen interessanter.

### 5.2 Wie funktioniert Imputing?

Als erstes muss festgestellt werden, welche verschiedenen Genvarianten in der Population vorkommen. Im Idealfall sind alle in einer Population vorkommenden Genvarianten bekannt. In der Praxis ist das aber leider nicht der Fall. In der Abbildung 8 ist beispielhaft ein Pedigree mit 15 Tieren mit allen vorkommenden Genvarianten dargestellt. Jedes Tier trägt eine väterliche und eine mütterliche Genvariante. Mittels statistischer Verfahren kann die Herkunft der Genvarianten, d.h. ob vom Vater oder Mutter vererbt, geschätzt werden.



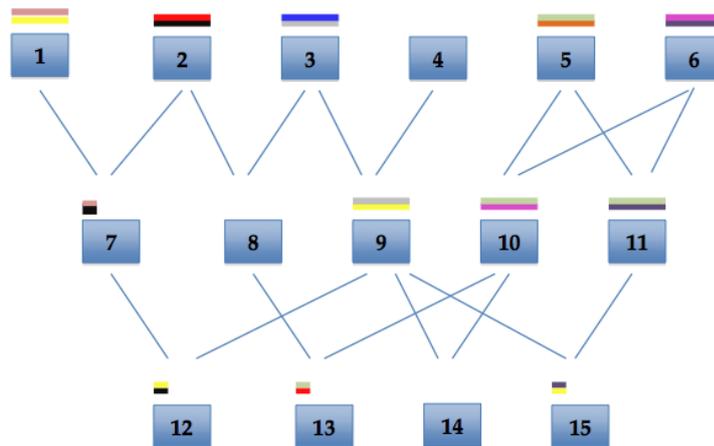
**Abbildung 8:** Pedigree mit den in der Population vorkommenden Genvarianten (nach Hickey, 2012)

Imputing kommt dann zur Anwendung, wenn nicht alle Tiere einer Population typisiert und daher auch nicht alle Genvarianten bekannt sind. So sind z.B. Tiere 1, 2, 3, 5, 6, 9, 10 und 11 mit dem Bovine50k-Chip typisiert (siehe Abbildung 9). Tiere 7, 12, 13 und 15 dagegen liegen nur LD-typisiert vor. Tiere 4, 8 und 14 sind gar nicht typisiert.

Zur Veranschaulichung ist der SNP-Genotypendatensatz der 15 Tiere in der Abbildung 10 dargestellt. Die fehlenden SNP können nun mittels Imputing hochgerechnet werden. Grundlage bilden die bekannten Genvarianten der 50k-typisierten Tiere. Wichtig ist, dass die vorkommenden Genvarianten in einer Population möglichst vollständig mit der 50k-typisierten Referenzpopulation abgedeckt werden. Die Referenzpopulation sollte daher wichtige, einflussreiche Ahnen mindestens 50k-typisiert enthalten.

Grundsätzlich lässt sich die Methodik des Imputing in zwei Gruppen einteilen: familien- und populationsbasiertes Imputing. Familienbasiertes Imputing folgt den Regeln der Mendelschen Vererbungslehre und nutzt Informationen von verwandten Tieren. Dabei stützt man sich auf Kopplungsinformation zwischen benachbarten SNP innerhalb einer Familie. Kopplung heisst, dass benachbarte SNP innerhalb einer Familie gehäuft gemeinsam vererbt werden. Dieses Verfahren ist sehr genau, wenn viele nahe verwandte Tiere des LD-typisierten Kandidatentieres 50k-typisiert sind.

Populationsbasiertes Imputing nutzt das Kopplungsungleichgewicht zwi-



**Abbildung 9:** Unterschiedlich typisierte Tiere im Pedigree (nach Hickey, 2012)

Tier1	AA	BB	AB	AB	BB	AA	AB	BB	AB	AA	AB	AA	AB	BB	AB	BB
Tier2	AB	AA	AB	BB	AA	AA	AB	BB	BB	BB	AB	AA	AB	BB	AA	AA
Tier3	AA	BB	AA	AA	BB	AB	BB	AB	AA	BB	AA	AA	AA	BB	AB	AB
Tier5	BB	AB	AA	AB	BB	AB	AA	AB	AA	BB	AB	AB	BB	AA	AB	BB
Tier6	AB	AA	BB	AB	AB	AA	BB	AB	BB	AB	BB	AA	AB	AB	AA	AB
Tier9	BB	BB	AA	AA	AB	BB	AB	AB	AA	AA	AB	BB	AB	AA	AB	BB
Tier10	AA	AA	AB	AB	BB	AB	BB	AB	AA	AB	BB	AB	AA	AB	BB	AB
Tier11	BB	AB	AA	AB	BB	AA	AA	AB	BB	AB	AA	AB	BB	AB	BB	AB
Tier7	.	.	.	BB	.	.	.	AB	.	.	.	AA	.	.	.	BB
Tier12	.	.	.	BB	.	.	.	AA	.	.	.	BB	.	.	.	BB
Tier13	.	.	.	AB	.	.	.	AA	.	.	.	AB	.	.	.	AA
Tier15	.	.	.	AA	.	.	.	BB	.	.	.	AB	.	.	.	AA
Tier4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Tier8	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Tier14	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

**Abbildung 10:** Beispiel für einen Genotypendatensatz mit 50k-, LD- und nicht genotypisierten Tieren

schen benachbarten SNP in der Population. Diese Methode ist speziell nützlich bei unverwandten Tieren oder wenn keine direkten Vorfahren eines Kandidaten typisiert sind. Um optimale Imputierungsgenauigkeiten zu erzielen, sind die beiden Ansätze in der Regel in Imputing-Programmen kombiniert.

Das Softwarepaket FImpute (Sargolzaei et al., 2012), das seit März 2013 routinemässig angewendet wird, nutzt sowohl familienbasiertes als auch populationsbasiertes Imputing.

## 5.3 Genauigkeit des Imputings

Qualitas AG arbeitete intensiv am Vergleich verschiedener Verfahren des Imputings vom LD-Chip auf den Bovine50k-Chip. Die Verfahren wurden an Datensätzen von Braunvieh, Holstein, Simmental und Swiss Fleckvieh getestet, wobei Braunvieh (BV) separat und Holstein, Simmental und Swiss Fleckvieh (MIX) gemeinsam analysiert wurden. Insgesamt standen bei BV 3'738 und bei MIX 4'743 Tiere zur Verfügung. Um die Genauigkeit des Imputings zu überprüfen, wurden jeweils die 20 % jüngsten Tiere selektiert. Bei ihnen wurden ausser den 9'000 LD Genotypen alle Genotypen der realen 50k-Typisierung auf „unbekannt“ gesetzt, um ein Imputing durchzuführen. Die Genauigkeit des Imputings wurde anhand der Übereinstimmung zwischen originalen und imputierten 50k-SNP festgestellt.

Die Tabelle 2 zeigt den durchschnittlichen Anteil an korrekt (% Korrekt) und inkorrekt (% Inkorrekt) imputierten Genotypen und die durchschnittliche Korrelation zwischen originalen und imputierten Genotypen für BV und MIX. Die Ergebnisse sind nach Verwandtschaftsgrad der Kandidaten zur 50k-genotypisierten Referenzpopulation dargestellt. Die Ergebnisse zeigen, dass unabhängig von der Rasse in mehr als 96% der Fälle das Imputing den korrekten Genotyp ermittelt hat. Daneben ist aber auch ersichtlich, dass die Genauigkeit des Imputings von der Verwandtschaft zwischen den LD-typisierten Kandidaten und der 50k-typisierten Referenzpopulation abhängt. Der höchste Anteil an korrekt imputierten Genotypen und die höchste Korrelation zwischen originalen und imputierten Genotypen ergeben sich für Tiere, bei denen beide Eltern 50k-typisiert sind. Der maximale Anteil an korrekt imputierten Genotypen liegt bei 99.7 %. Im Durchschnitt werden 98.6 % für BV und 97.7 % für MIX der Genotypen korrekt imputiert. Die mittlere Korrelation liegt zwischen 0.95 und 0.98. Die Fehlerraten reichen von 1.4 bis 3.2 % und steigen mit abnehmender Verwandtschaft an.

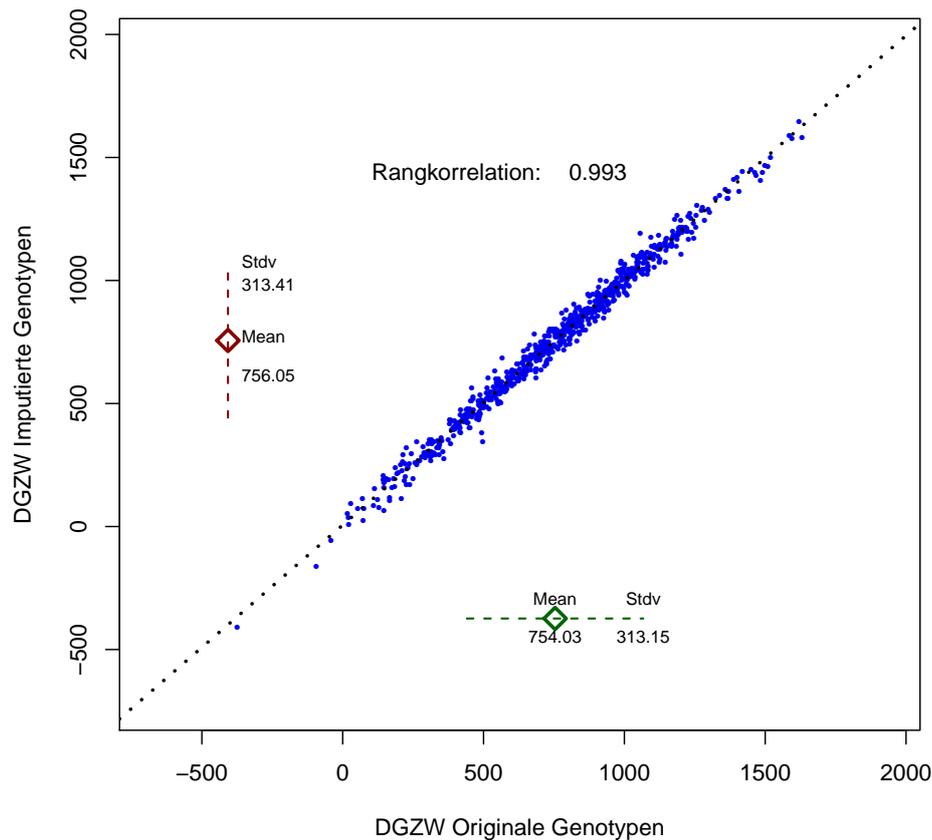
**Tabelle 2:** Durchschnittlich % korrekt und % inkorrekt imputierte Genotypen und durchschnittliche Korrelation zwischen imputierten und originalen Genotypen (MGV = mütterlicher Grossvater)

Gruppe	BV			
	Anzahl	%Korrekt	%Inkorrekt	Korrelation
<i>Beide Eltern</i>	27	98.6	1.4	0.98
<i>Vater und MGV</i>	573	98.1	1.9	0.97
<i>Vater</i>	55	97.6	2.4	0.96
<i>Andere</i>	68	97.3	2.7	0.95
	MIX			
<i>Vater und MGV</i>	331	97.7	2.3	0.96
<i>Vater</i>	59	96.8	3.2	0.95
<i>Andere</i>	544	96.9	3.1	0.95

## 5.4 Genomische Zuchtwertschätzung mit imputierten Genotypen

Um die Auswirkungen des Imputings auf die DGZW zu überprüfen, wurden DGZW für die Kandidaten sowohl ausgehend von originalen, als auch ausgehend von imputierten Genotypen für alle Merkmale der Routinezuchtwertschätzung berechnet und miteinander verglichen. Unterschiede in der Rangierung der Kandidaten nach DGZW mit originalen und imputierten Genotypen lassen sich durch Berechnung der Rangkorrelation feststellen. Rangkorrelationen für alle Merkmale sind sowohl bei BV als auch bei MIX sehr hoch und liegen im Bereich von 0.989 und 0.999. Die hohen Rangkorrelationen zeigen, dass die selben Selektionsentscheide getroffen werden, egal ob reale oder imputierte Genotypen dem DGZW zu Grunde liegen. Zu beachten ist, dass es trotz der hohen Rangkorrelationen im Einzelfall zu Abweichungen zwischen den DGZW basierend auf originalen und basierend auf imputierten Genotypen kommen kann. Die Abbildung 11 stellt den DGZW basierend auf originalen und imputierten Genotypen für BV für das Merkmal Milchmenge dar. Dieser Abbildung kann entnommen werden, dass nicht alle Kandidaten

vollkommen auf einer Linie liegen, d.h. in einzelnen Fällen unterscheiden sich die beiden DGZW mehr oder weniger stark. Da nicht alle SNP korrekt imputiert werden, ist weiter damit zu rechnen, dass die Sicherheit der DGZW basierend auf imputierten Genotypen in Abhängigkeit von der Genauigkeit des Imputings leicht vermindert ist.



**Abbildung 11:** XY-Diagramm von DGZW für Milchmenge bei BV basierend auf originalen und imputierten Genotypen

## 6 Aktuelle Forschungsarbeiten

### 6.1 Imputing vom 50k-Chip auf den BovineHD-Chip

Aktuelle Forschungsarbeiten im Bereich der genomischen Selektion beschäftigen sich u.a. mit der Nutzung verschiedener SNP-Chips und sogar der vollständigen Genomsequenz. Neben dem LD- und Bovine50k-Chip bietet die Firma Illumina einen HD-Chip mit 777'0000 SNP an. Auf den BovineHD-Chip und die Sequenzdaten werden besonders Hoffnungen im Bereich der rasenübergreifenden genomischen Zuchtwertschätzung gesetzt. Davon könnten vor allem kleine Populationen profitieren. Qualitas AG hat in einer ersten Studie die Genauigkeit des Imputings vom Bovine50k-Chip auf den BovineHD Chip untersucht.

Für die Untersuchung standen 6'106 50k und 880 HD Genotypen von Stieren und Kühen der Rassen Original Braunvieh, Braunvieh und Brown Swiss zur Verfügung. Das Imputing erfolgte von 39'004 auf 627'306 SNP. Um die Genauigkeit des Imputings zu überprüfen, wurden die HD Genotypen von 365 Stieren und Kühen geboren zwischen 2004 und 2008 auf fehlend gesetzt. Die Genauigkeit des Imputings wurde überprüft, indem die originalen mit den imputierten Genotypen der 365 Tiere verglichen wurden. Berechnet wurden die quadrierte Korrelation ( $R^2$ ) zwischen den originalen und imputierten Genotypen und der Prozentsatz an korrekt imputierten Genotypen. In Tabelle 3 sind die Ergebnisse in Abhängigkeit des Verwandtschaftsgrades zwischen den 365 imputierten Tieren und der restlichen HD- und 50k-genotypisierten Population dargestellt. Die Ergebnisse zeigen für die 365 Stiere und Kühe eine sehr gute Übereinstimmung zwischen den originalen und imputierten Genotypen. Die durchschnittlichen  $R^2$  liegen zwischen 0.94 und 0.98. Die Genauigkeit des Imputings ist abhängig von der Verwandtschaft zwischen den 50k-genotypisierten Kandidaten und der HD-genotypisierten Referenzpopulation. So sind die  $R^2$  für Tiere bei denen z.B. beide Eltern, Vater und mütterlicher Grossvater, Mutter und väterlicher Grossvater HD-typisiert vorliegen, am höchsten. Bei den restlichen Gruppen ist das  $R^2$  niedriger. Der maximale Anteil an korrekt imputierten Genotypen liegt bei 99.58%.

**Tabelle 3:** Durchschnittliche % Korrekt imputierter Genotypen und durchschnittliche quadrierte Korrelation ( $R^2$ ) zwischen originalen und imputierten Genotypen (MGV = mütterlicher Grossvater, VGV = väterlicher Grossvater)

Gruppe	N	$R^2$	%korrekt
<i>Beide Eltern HD</i>	57	0.98	99.58
<i>Vater + MGV HD</i>	79	0.98	99.25
<i>Mutter + VGV HD</i>	4	0.98	99.42
<i>Vater HD</i>	75	0.97	98.95
<i>Mutter HD</i>	6	0.94	97.85
<i>Vater HD + Mutter 50k</i>	11	0.98	99.35
<i>Vater 50k + Mutter HD</i>	25	0.97	98.74
<i>Vater 50k + MGV 50k</i>	26	0.95	98.37
<i>Vater 50k</i>	39	0.96	98.55
<i>Andere</i>	43	0.94	97.97

Bei abnehmender Verwandtschaft nimmt der Anteil an korrekt imputierten Genotypen ab.

## 6.2 Genomische Zuchtwerte für kleine Rassen

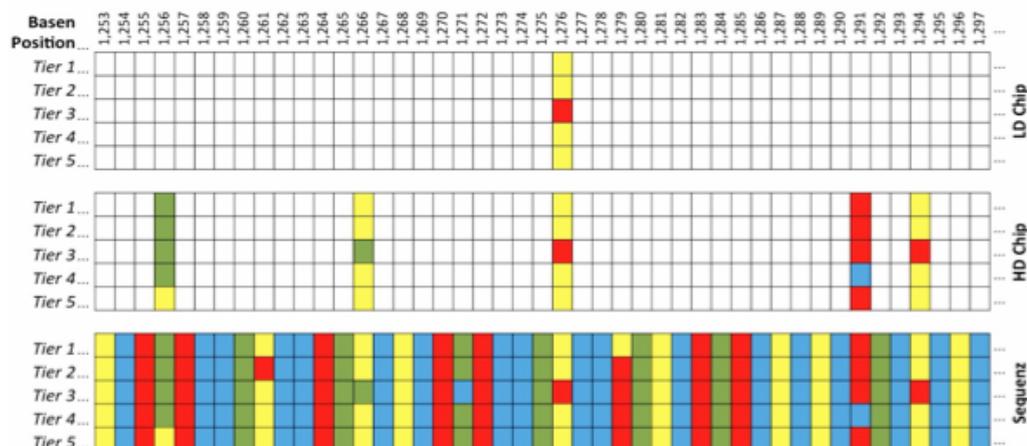
Da die Grösse des Trainingsdatensatzes der entscheidende Faktor für ausreichend genaue SNP-Effekte ist, haben Rassen mit wenig nachzuchtgeprüften Stieren, wie z.B. Simmental oder Original Braunvieh, bisher über keine genomische Zucht verfügt. Generell gilt, je grösser der Trainingsdatensatz, desto besser. Die Genotypisierung mit dem BovineHD Chip zusammen mit rassenübergreifenden Effektschätzungen können für kleine Rasse eine Lösung darstellen. Im Rahmen von internationalen und nationalen Projekten werden Stiere der Rassen Simmental und Original Braunvieh genotypisiert. Resultate aus ersten Effektschätzungen liegen vor und sehen vielversprechend aus.

## 6.3 Nutzung von Sequenzdaten in der genomischen Zuchtwertschätzung

Sequenzdaten stellen sozusagen den ultimativen SNP-Chip dar, da die für die Ausprägung eines Merkmals verantwortliche Genvariante in fast jedem Fall im Datensatz enthalten ist. Die SNP-Effekte basierend auf Sequenzdaten sollten somit die tatsächlichen Geneffekte deutlicher darstellen. Dadurch ist eine Erhöhung der Sicherheit der genomischen Zuchtwerte zu erwarten. Zudem könnte die genomische Zuchtwertschätzung nachhaltiger werden, d.h. die geschätzten SNP-Effekte bleiben über Generationen konstanter. Die Nutzung von Sequenzdaten in der genomischen Zuchtwertschätzung bietet auch Vorteile für die rassenübergreifende genomische Zuchtwertschätzung. Gerade in Merkmalen mit niedriger Erbllichkeit erhofft man sich die deutlichsten Verbesserungen.

Aus der Sicht eines Genetikers wäre es ideal, alle Tiere in der Population zu sequenzieren. Aus finanzieller Sicht ist dies aber aktuell nicht möglich. Eine bezahlbare Alternative ist es, nur wichtige Schlüsseltiere zu sequenzieren. Werden die wichtigsten Stiere analysiert, können die Sequenzgenotypen für alle derzeit genotypisierten Tiere hochgerechnet werden. Die SNP-Genotypen der Sequenz können dann in der aktuell 50k- oder HD-typisierten Population imputiert werden. Mit dieser Vorgehensweise könnte die genomische Zuchtwertschätzung auf etwa 20 Millionen SNP anstatt der bisher 50'000 SNP Genotypen basieren. Die Qualitas AG ist Partner in zwei internationalen Projekten (Gene2Farm, SLIG), in denen Sequenzinformationen von Schlüsseltieren der Schweizerischen Milchviehpopulationen analysiert werden.

Im Gene2Farm Projekt arbeitet die Qualitas AG im Auftrag von Braunvieh Schweiz und Swissherdbook zusammen mit 19 europäischen Forschungs- und Industriepartnern an der Sequenzierung von Schlüsseltieren der wichtigsten Rinderrassen Europas. Neben der Erstellung einer umfassenden Sequenzdatenbank sollen detaillierte Analysen von Genomstruktur und Gemeinsamkeiten innerhalb und zwischen den Rassen gemacht werden. Zusammen mit Zuchtorganisationen werden auch neue Methoden zur Leistungskontrolle entwickelt. Das Projekt hat eine Laufzeit von vier Jahren und wird von der Europäischen Union gefördert.



**Abbildung 12:** Übersicht vom LD Chip zur Sequenz. Je ein Buchstabe des genetischen Alphabets wird mit unterschiedlichen Farben (A gelb, C rot, G blau und T grün) dargestellt. Die Abbildung zeigt die unterschiedliche SNP-Dichte, vom kleinen LD-Chip mit nur wenig SNP bis hin zur Sequenz, die alle SNP beinhaltet. Tier 3 beispielsweise ist an Position 1271 „blau“ (Buchstabe „G“), wobei alle anderen Tiere grün („T“) sind. Diese Variation ist weder auf den 50k- noch auf dem LD-Chip sichtbar. Sind die wichtigsten Stiere einer Rasse sequenziert, können die fehlenden Positionen (weisse Kästchen) des 50K bzw. des LD Chip durch Imputing aufgefüllt werden

In Zusammenarbeit mit mehreren Schweizer Forschungs- und Industriepartnern (agn Genetics, Hochschule für Agrar-, Forst- und Lebensmittelwissenschaften HAFL und swissgenetics) und einer amerikanischen Universität (Iowa State University) startete die Qualitas AG ein zweites Sequenzierungsprojekt (SLIG). Hauptziel des SLIG Projekts ist es, die Genauigkeit von genomischen Zuchtwerten mit Sequenzdaten zu erhöhen. Weiter soll im Rahmen des Projektes ein Werkzeug zur Inzuchtkontrolle und Erhaltung der genetischen Vielfalt entwickelt werden. Das SLIG Projekt wird von der Arbeitsgemeinschaft Schweizerischer Rinderzüchter und der Kommission für Technologie und Innovation gefördert und hat eine Laufzeit von drei Jahren.

## 6.4 Nutzung von typisierten Kühen in der genomischen Selektion

Die Genauigkeit der genomischen Zuchtwerte ist abhängig von der Grösse, der Zusammensetzung und der Qualität des Trainingsdatensatzes. Er ist sozusagen das Herzstück der genomischen Zuchtwertschätzung.

Das Potenzial von typisierten Stieren mit sicheren ZW (hohes B%) für den Trainingsdatensatz ist bei kleineren Rassen beschränkt. Auch internationale Austauschaktivitäten oder länderübergreifende Projekte (z.B. Intergenomics von Interbull beim Braunvieh) können dieser Tatsache nichts entgegenhalten. Der Einbezug von genotypisierten Kühen, die sichere Zuchtwerte aufweisen, ist ein vielversprechender Lösungsansatz. Dieser wird im Ausland zum Teil bereits angewendet. Da Kühe aber nie die Bestimmtheitsmasse von KB-Stieren erreichen, müssen viel mehr weibliche Tiere typisiert werden. Als Faustzahl kann folgendes angenommen werden: 7 Kühe ersetzen einen Stier, abhängig von der Erblichkeit des Merkmals.

Braunvieh Schweiz besitzt etwa 1300 weibliche Genotypen, die für die Effektschätzung infrage kommen. In Testläufen wurden diese in den Trainingsdatensatz integriert. Die Resultate zeigen, dass dies noch zu wenige Kühe sind, um Verbesserungen bei der Sicherheit der genomischen Zuchtwerte erzielen zu können. Zudem ist die genetische Breite durch die grossen Halbschwistergruppen und durch eine Vorselektion durch das Elitekuhprogramm zu wenig abgedeckt.

Idealerweise müssten ab einer bestimmten B%-Grenze und abhängig vom Pedigree zufällig ausgewählte Kühe – egal, ob sie gute oder schlechte Zuchtwerte haben – in grosser Anzahl typisiert werden. Da die finanziellen Ressourcen immer der limitierende Faktor sind, wird intensiv daran gearbeitet, wie die Kühe für die Typisierung optimal ausgewählt werden können.

Ein interessanter Zusatznutzen könnte sein, dass Merkmale, die teuer in der Erfassung sind und deshalb nicht in der ganzen Population erhoben werden können, gezielt an einer typisierten Kuhpopulation erfasst werden. Spezifische Gesundheitsmerkmale, Effizienzmerkmale, Methanausstoss oder Klauenscoring könnten auf diese Art und Weise züchterisch bearbeitet werden.

Version 20. Februar 2014

QUALITAS<sup>+</sup>

Chamerstrasse 56

CH-6300 Zug

Tel.: +41 (0)41 768 92 92

Fax.: +41 (0)41 768 92 99

[info@qualitasag.ch](mailto:info@qualitasag.ch)

[www.qualitasag.ch](http://www.qualitasag.ch)