Structure-Preserving GANs

Jeremiah Birrell (UMass), Markos A. Katsoulakis (UMass),
Luc Rey-Bellet (UMass), Wei Zhu (UMass)

ICML 2022

Generative adversarial networks (GANs)

- GANs use a pair of neural networks to learn a probability distribution.

Generative adversarial networks (GANs)

- GANs use a pair of neural networks to learn a probability distribution.
- **Zero-sum game** between discriminator and generator—**“the players”.**

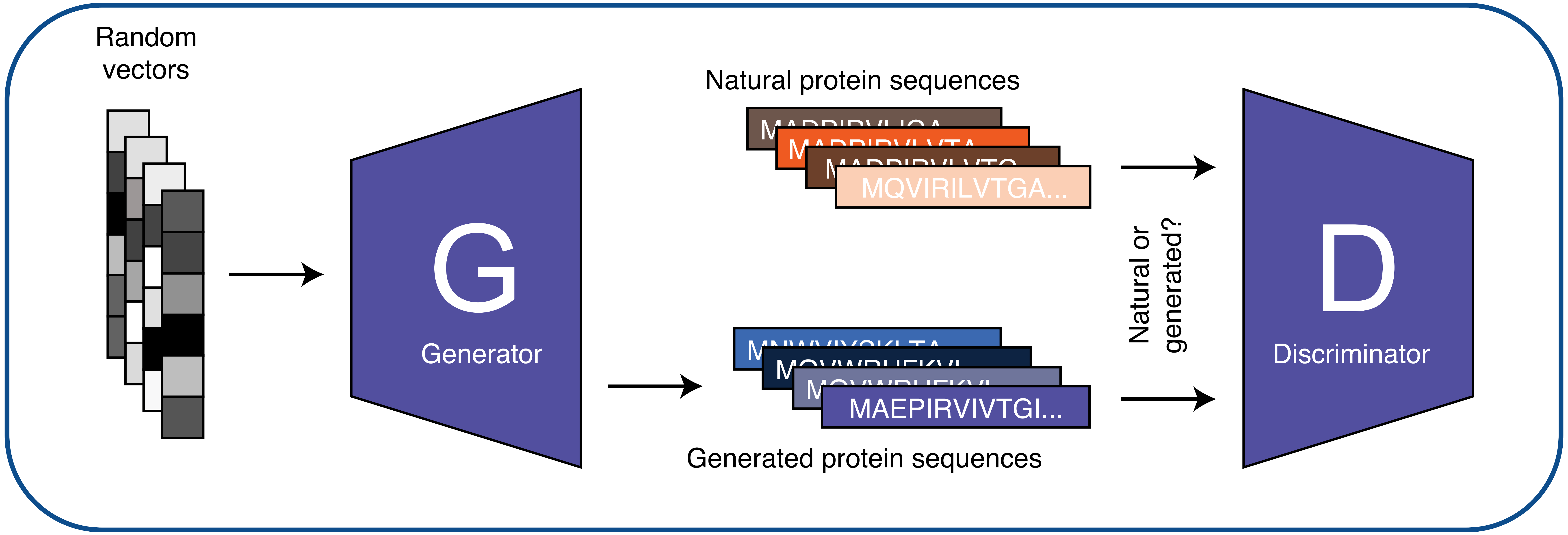


Figure: Repecka et al., *Nature Machine Intelligence* 2021

Generative adversarial networks (GANs)

- GANs use a pair of neural networks to learn a probability distribution.
- **Zero-sum game** between discriminator and generator—**“the players”.**
- Game ends when the players reach *consensus*: “fake data” looks like the “real” data.

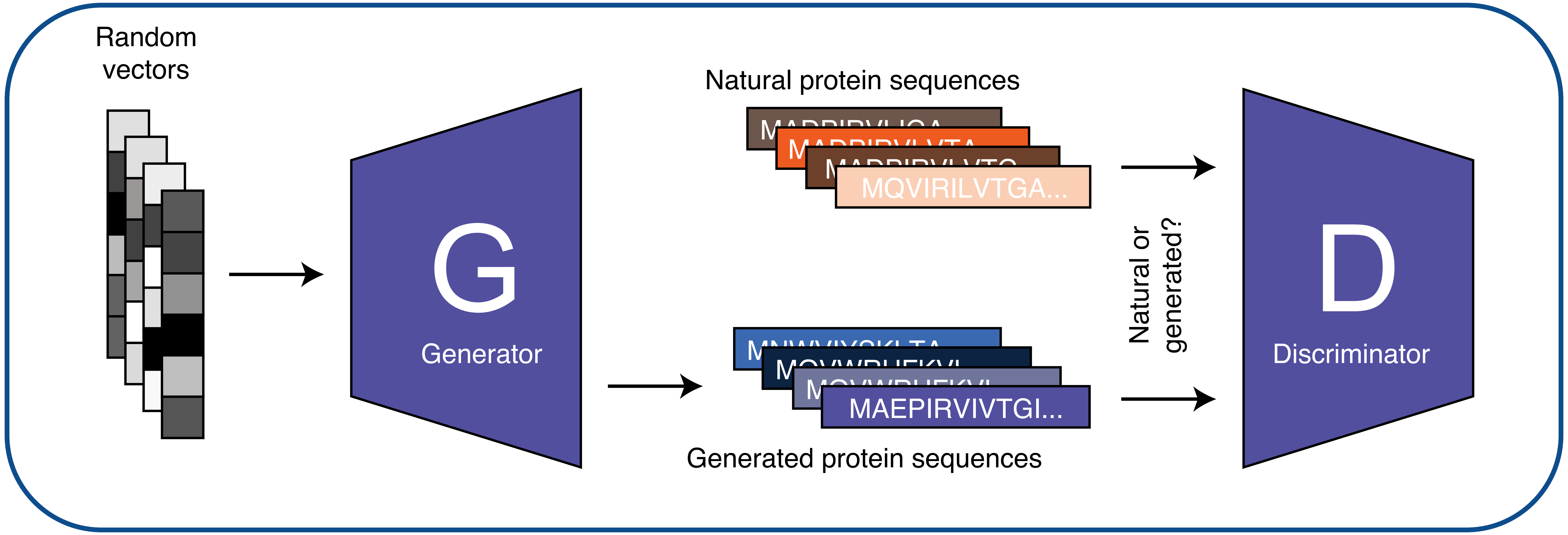
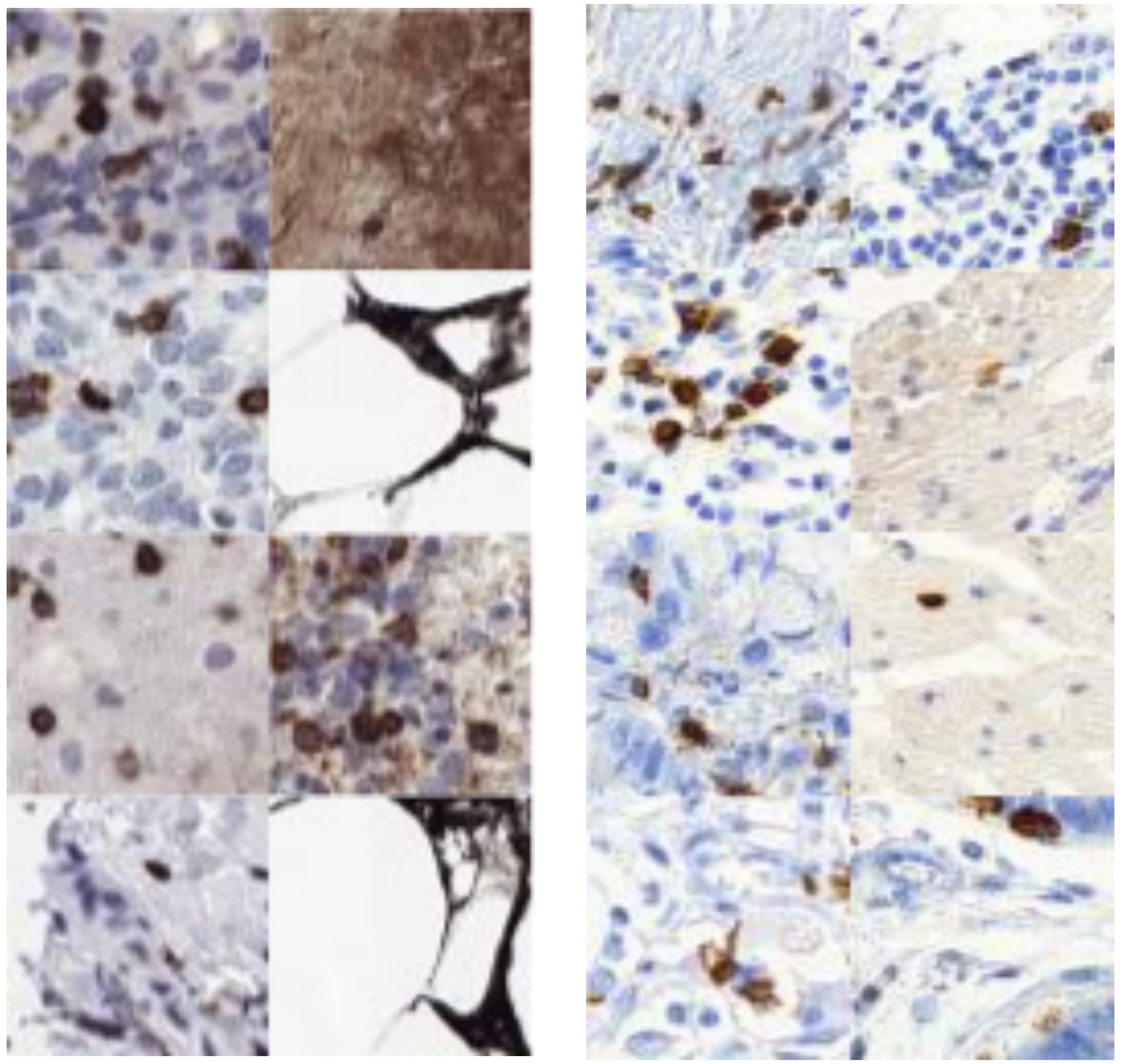


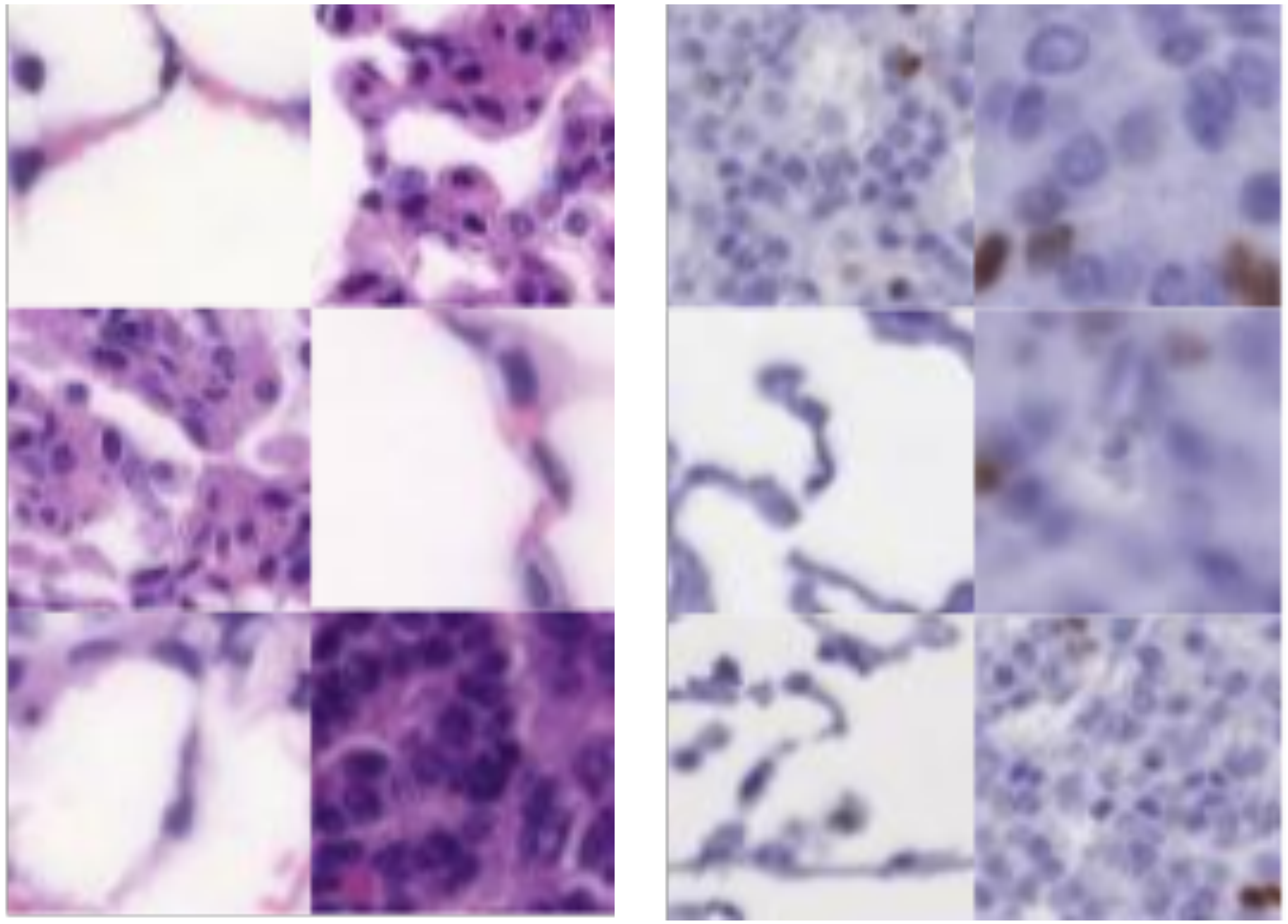
Figure: Repecka et al., *Nature Machine Intelligence* 2021

Structured target data & distribution Q

Q



LYSTO[1]



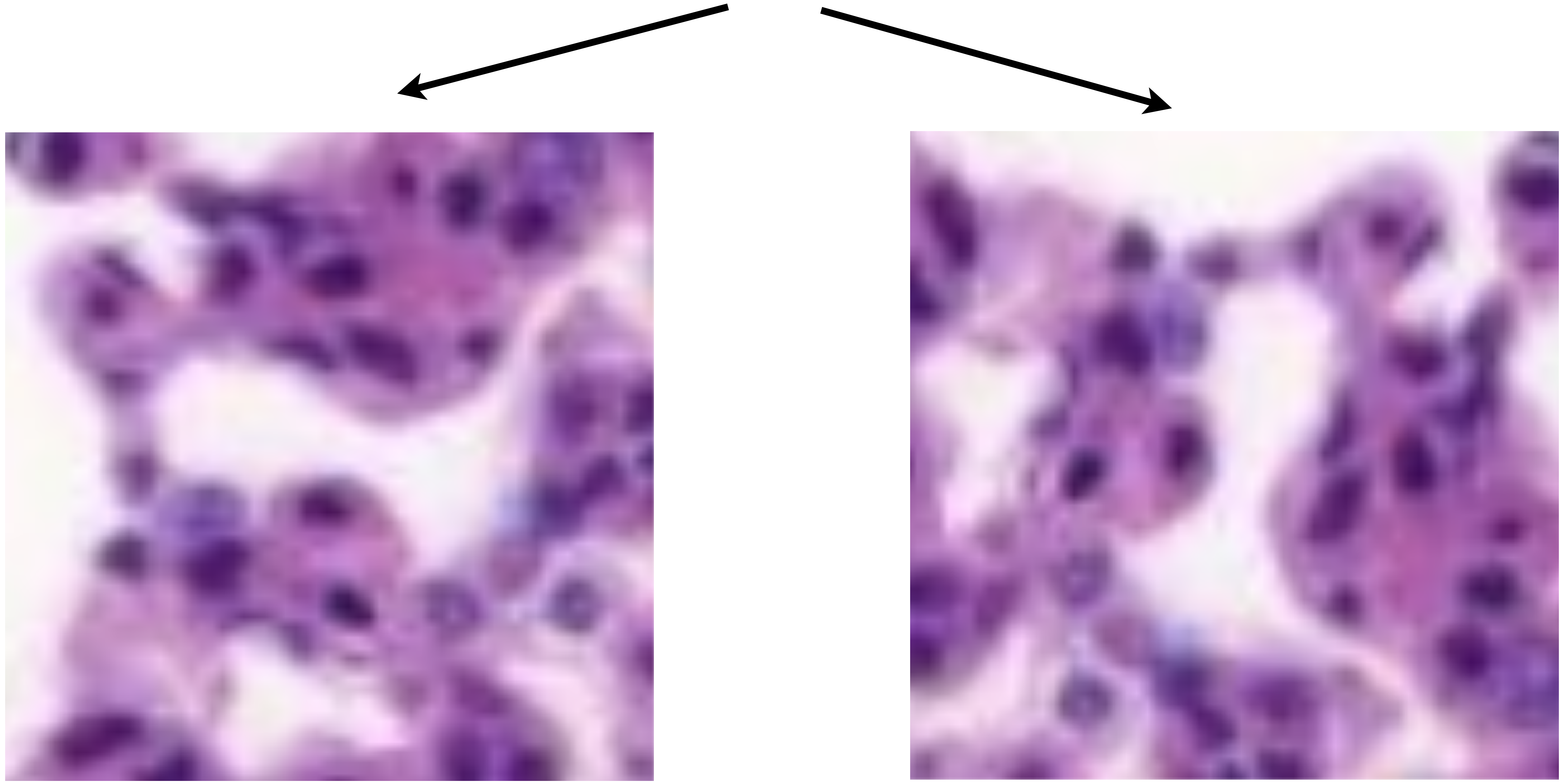
ANHIR[2]

1. *Ciompi et al., Zenodo 2019*
2. *Borovec et al., IEEE Transactions on Medical Imaging 2020*

Structured target data & distribution Q

equiprobable

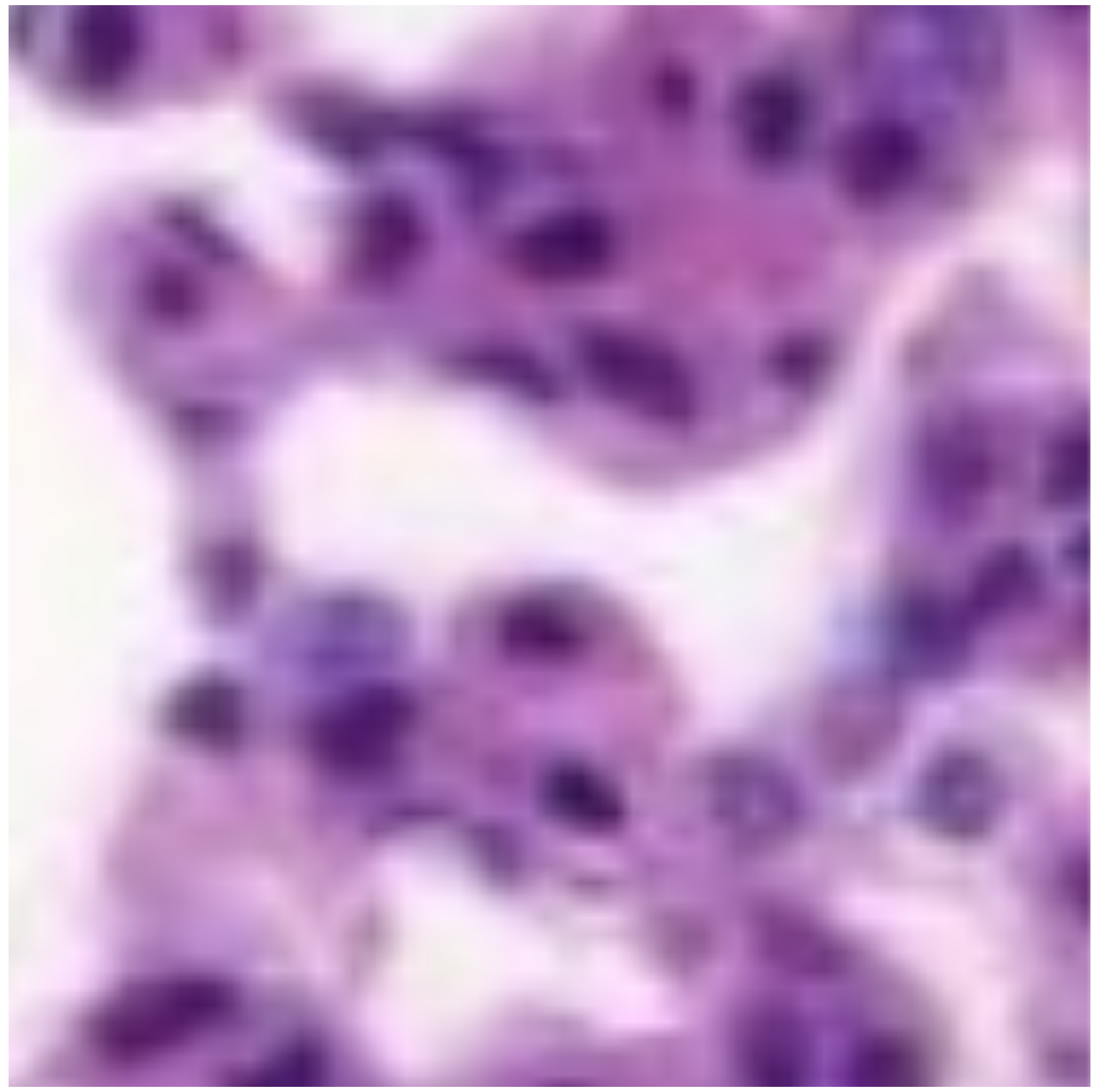
Q

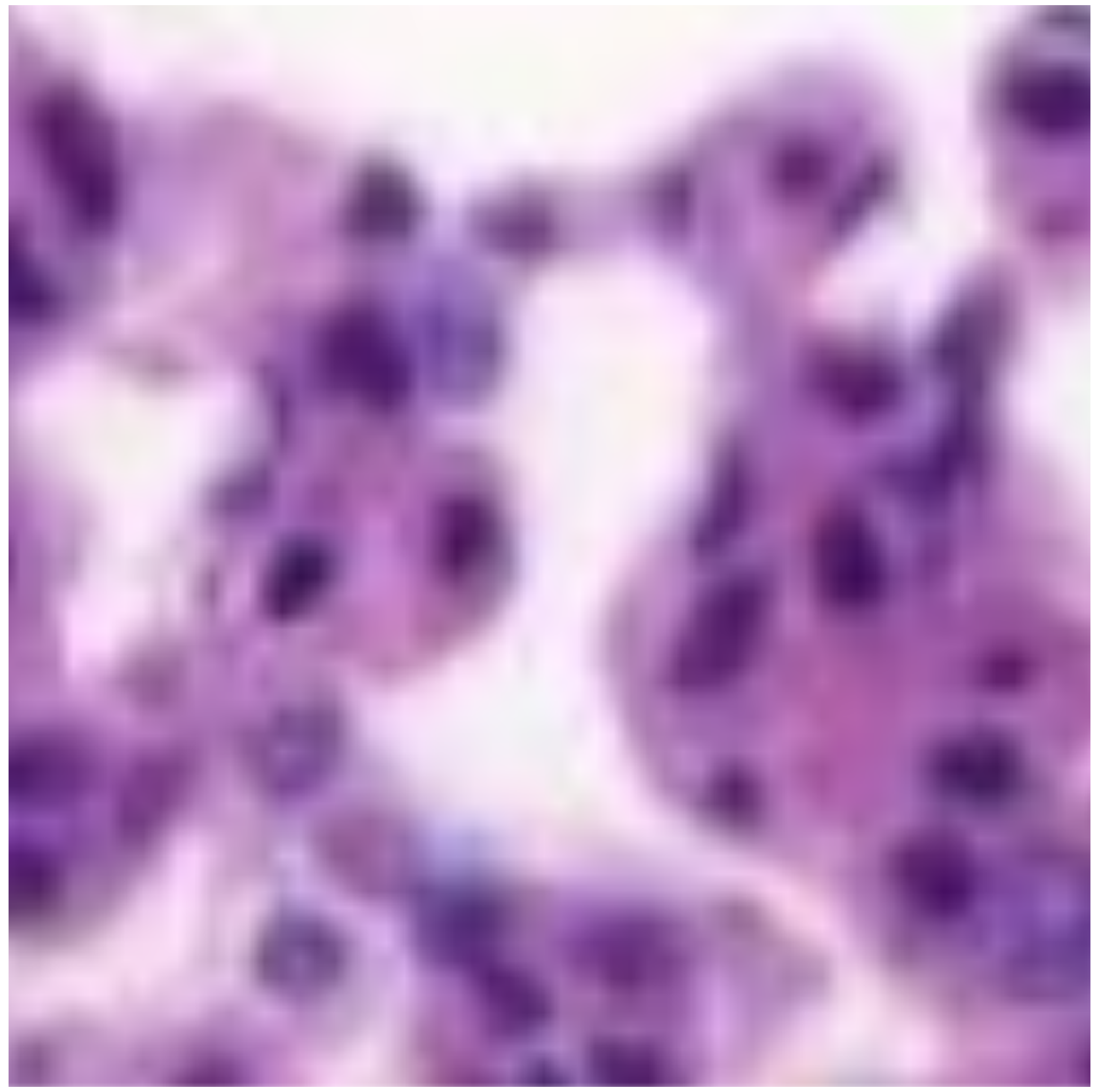


Structured target data & distribution Q

equiprobable

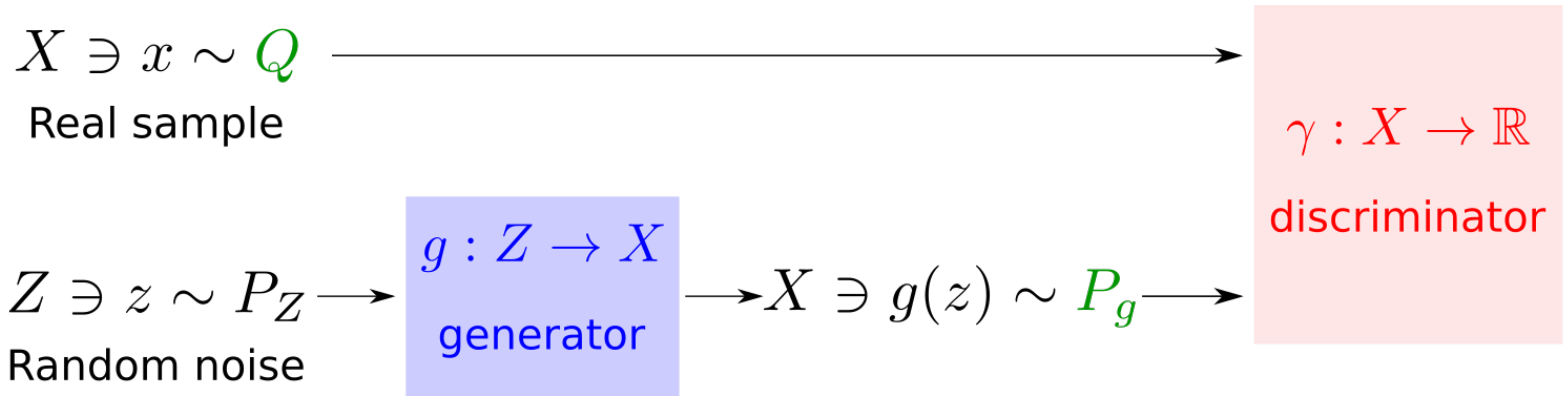
Q





Question: how to build **embedded structure** into GAN players (generators and discriminators) for data-efficient distribution learning?

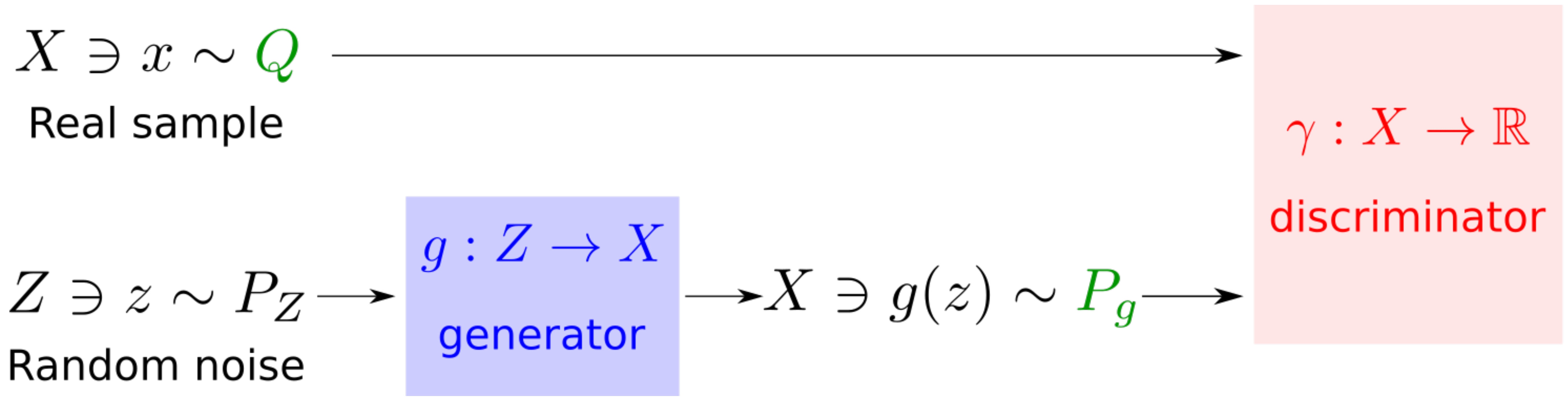
GAN is “probability distance” minimization



Mathematically, GANs can be formulated as minimizing some **variational divergence**, $D^{\Gamma}(Q \| P_g)$, between Q and P_g.

$$\min_{g \in G} D^{\Gamma}(Q \| P_g)$$

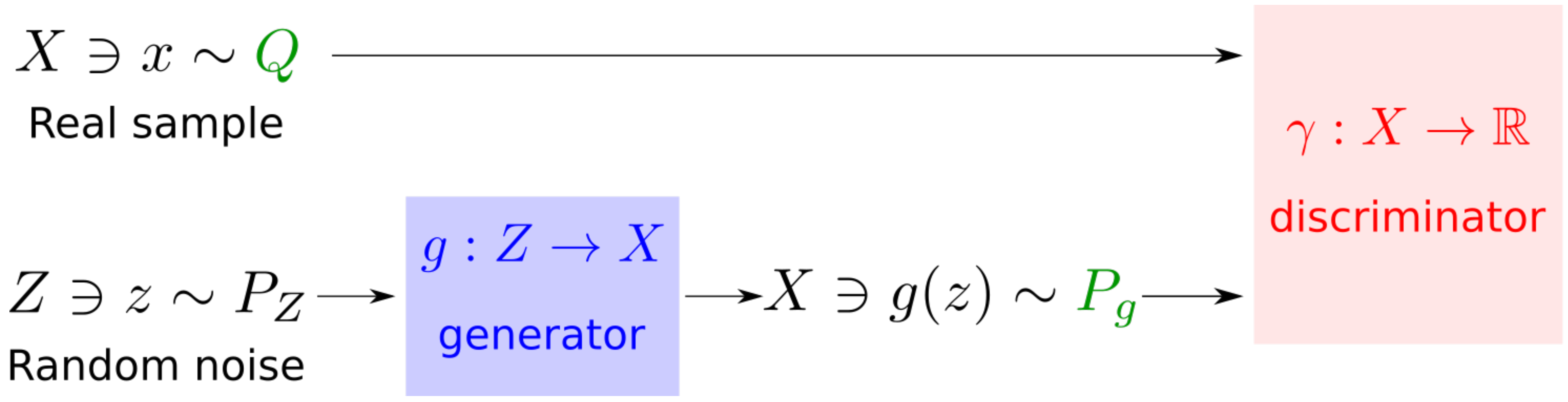
GAN is “probability distance” minimization



Mathematically, GANs can be formulated as minimizing some **variational divergence**, $D^{\Gamma}(Q \| P_g)$, between Q and P_g. “Distance” is determined by discriminators $\gamma \in \Gamma$.

$$\min_{g \in G} D^{\Gamma}(Q \| P_g) = \min_{g \in G} \max_{\gamma \in \Gamma} H(\gamma; Q, P_g)$$

GAN is "probability distance" minimization

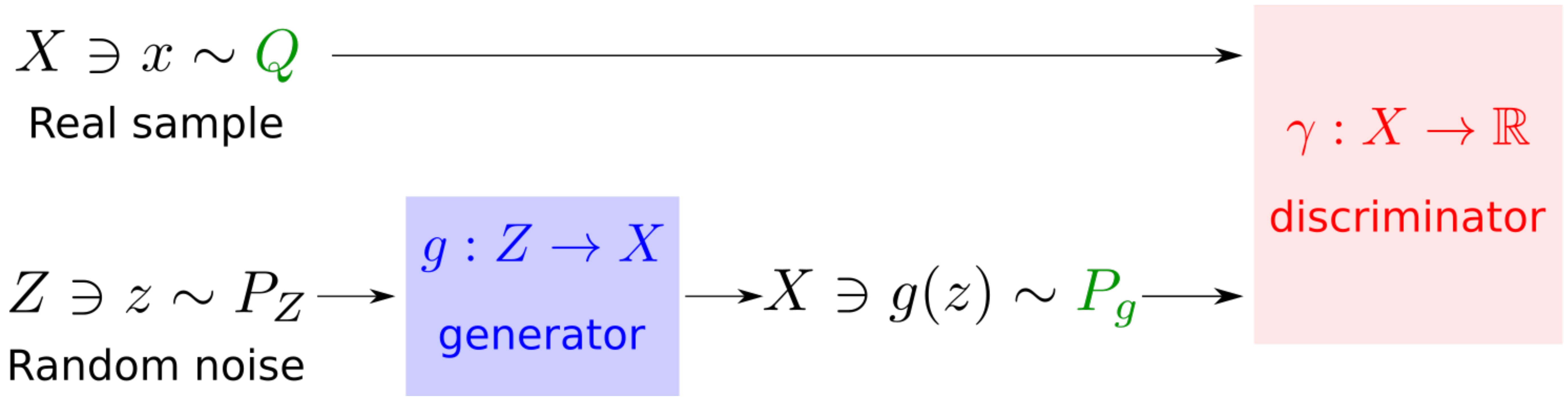


Mathematically, GANs can be formulated as minimizing some **variational divergence**, $D^{\Gamma}(Q\|P_g)$, between Q and P_g. "Distance" is determined by discriminators $\gamma \in \Gamma$.

$$\min_{g\in G} D^{\Gamma}(Q\|P_g) = \min_{g\in G}\max_{\gamma\in\Gamma} H(\gamma; Q, P_g)$$

- The original GAN [Goodfellow et al., 2014]: $\min_{g\in G}\max_{\gamma\in\Gamma} E_Q[\log\gamma] + E_{P_g}[\log(1-\gamma)]$

GAN is “probability distance” minimization

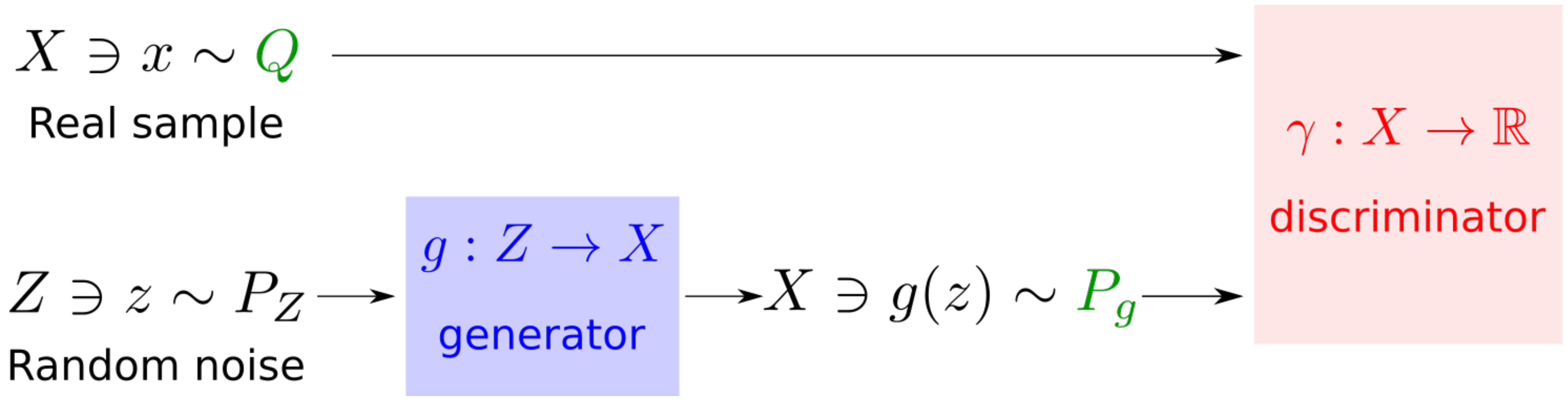


Mathematically, GANs can be formulated as minimizing some **variational divergence**, $D^\Gamma(Q \| P_g)$, between Q and P_g. “Distance” is determined by discriminators $\gamma \in \Gamma$.

$$\min_{g \in G} D^\Gamma(Q \| P_g) = \min_{g \in G} \max_{\gamma \in \Gamma} H(\gamma; Q, P_g)$$

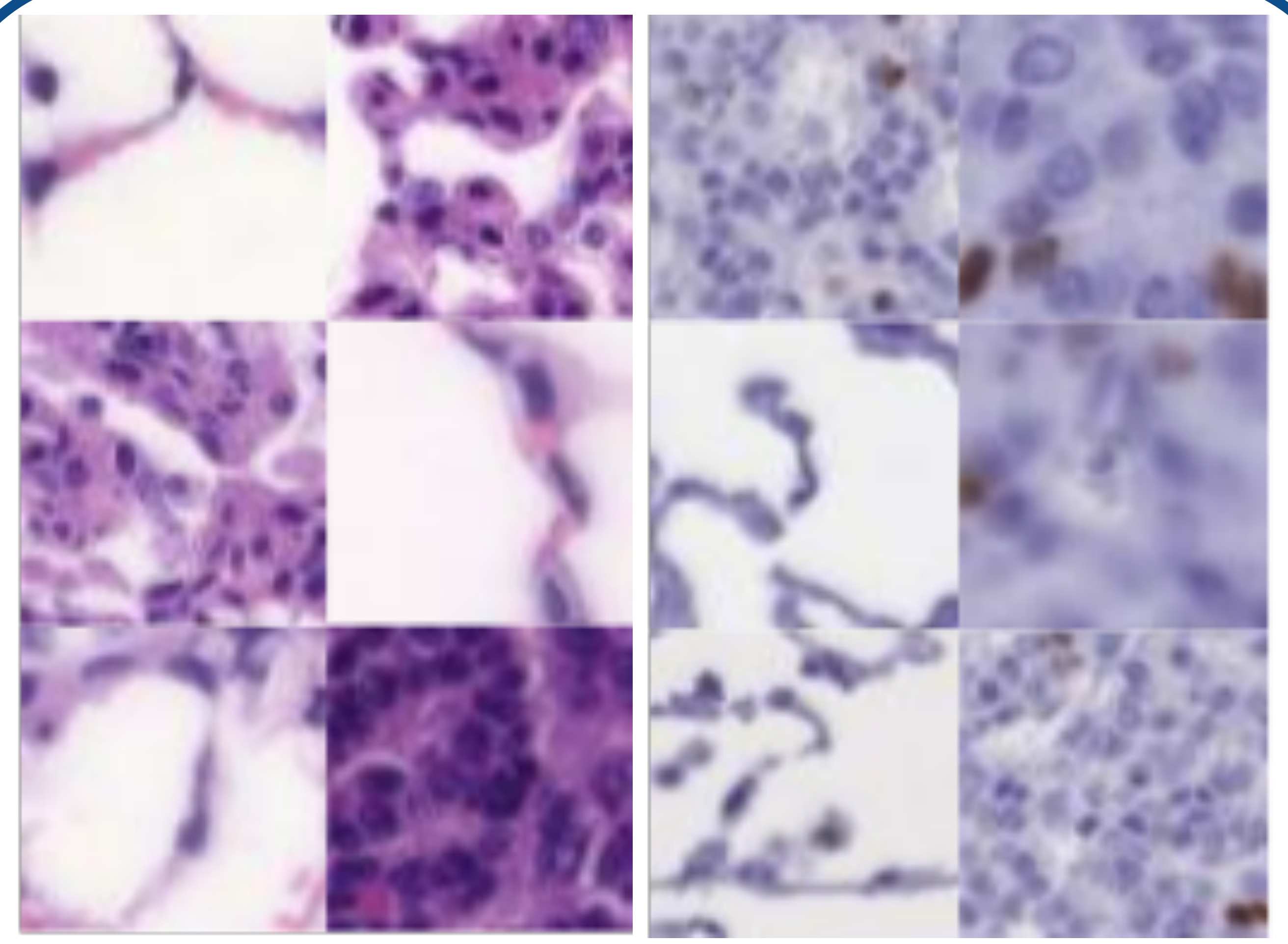
- The original GAN [Goodfellow et al., 2014]: $\min_{g \in G} \max_{\gamma \in \Gamma} E_Q[\log \gamma] + E_{P_g}[\log(1 - \gamma)]$
- Many other variational divergences: KL, f-divergences, Wasserstein, MMD, Sinkhorn...

GAN with embedded structure

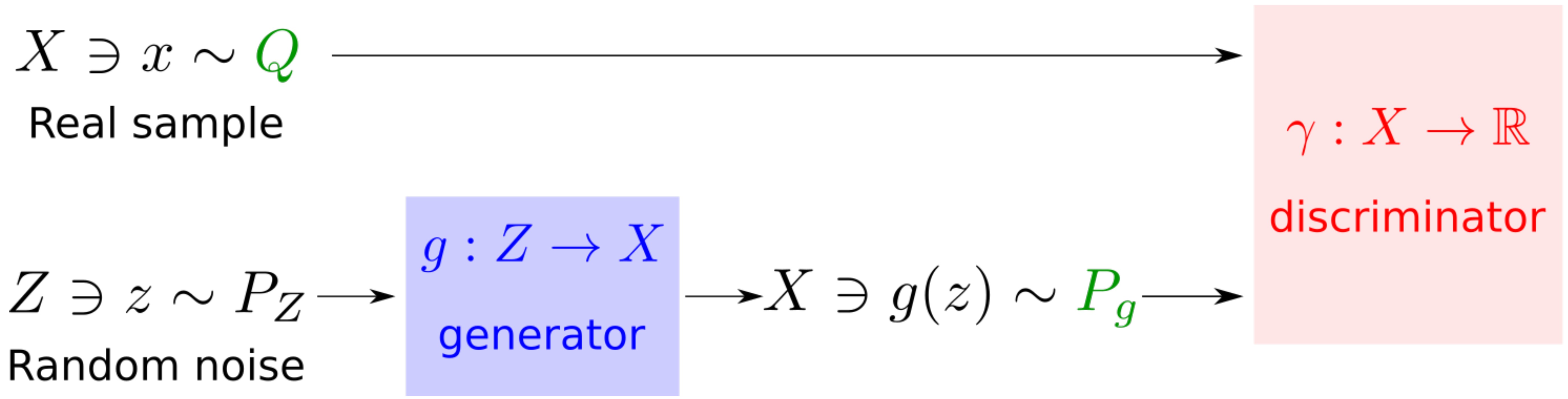


$$\min_{g\in G} D^{\Gamma}(Q\|P_g) = \min_{g\in G}\max_{\gamma\in\Gamma} H(\gamma; Q, P_g),$$

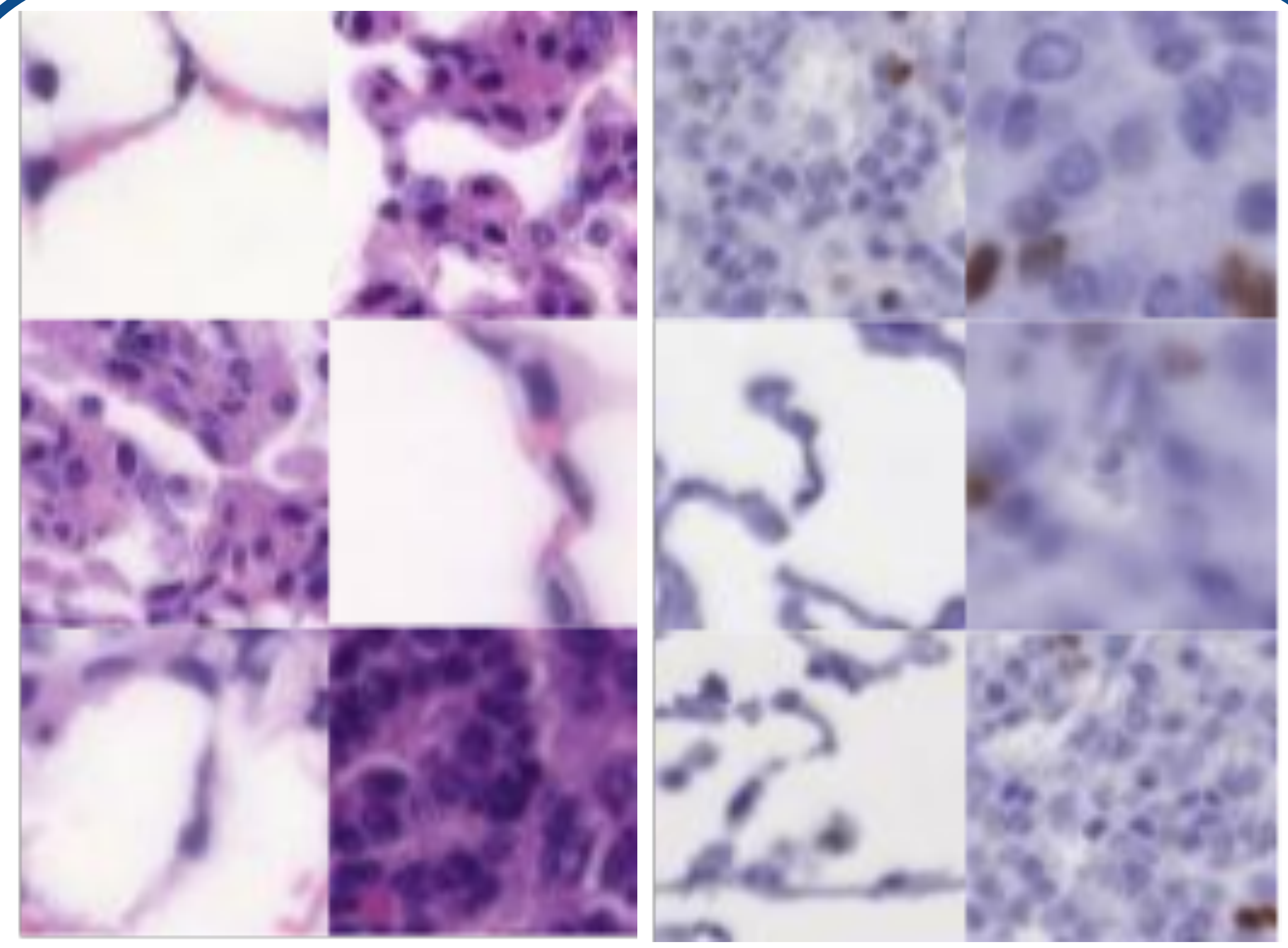
Q is Σ-invariant



GAN with embedded structure

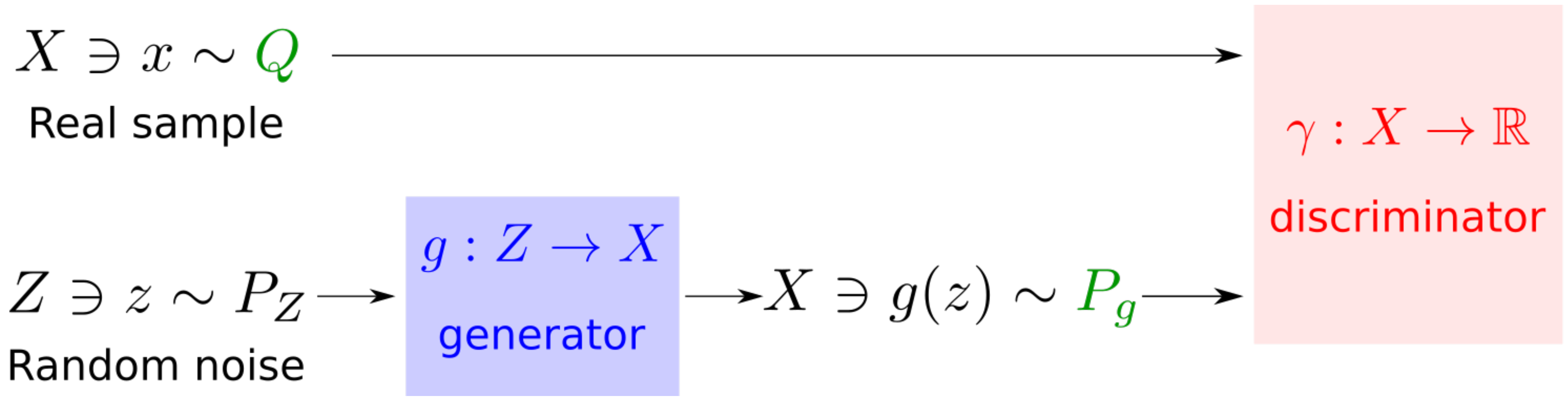


$$\min_{g \in G} D^{\Gamma}(Q \| P_g) = \min_{g \in G} \max_{\gamma \in \Gamma} H(\gamma; Q, P_g), \quad Q \text{ is } \Sigma\text{-invariant}$$



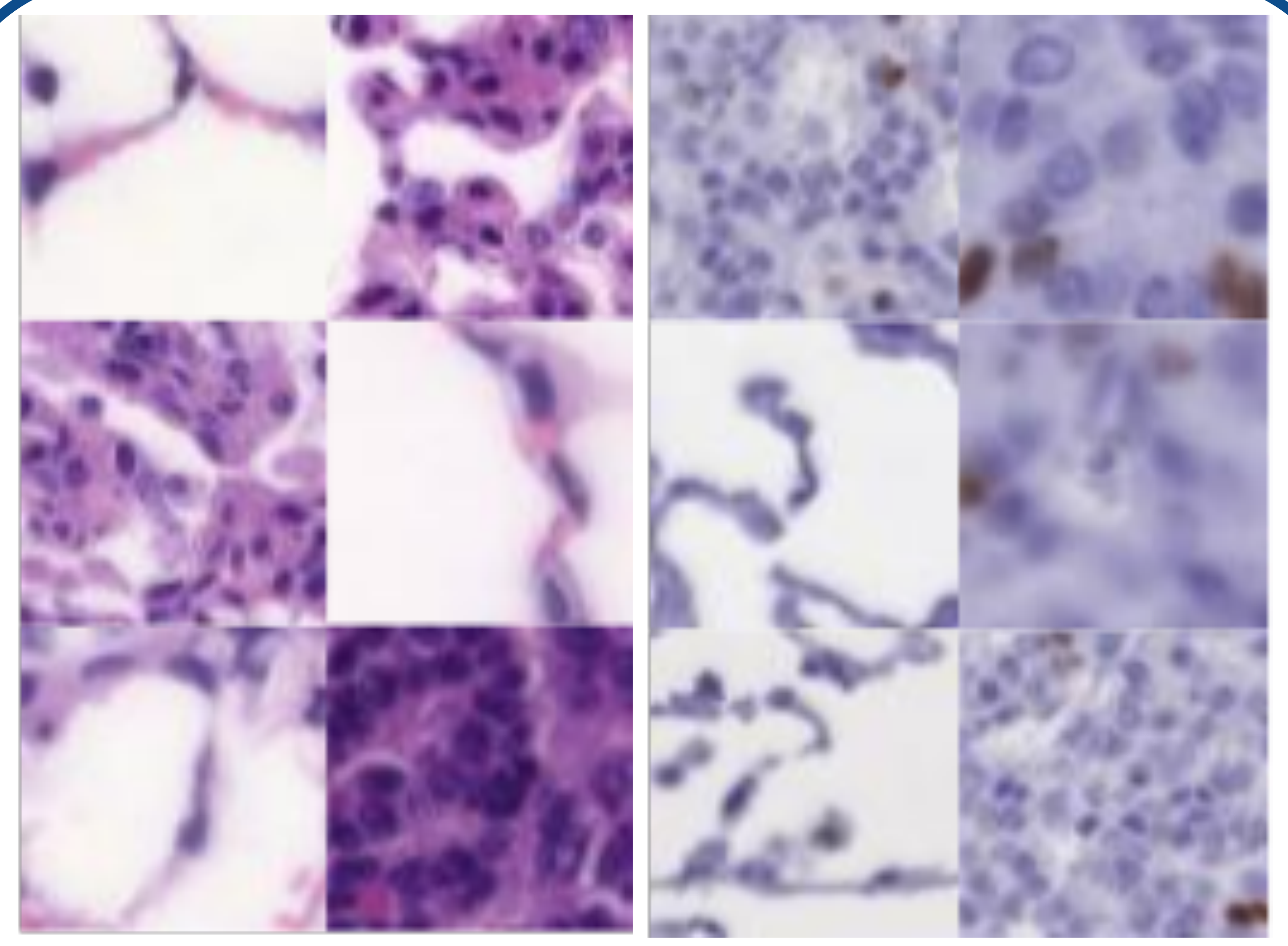
- Target distribution Q is invariant under a group Σ.
- Σ: rotation, translation, roto-reflection, permutation, etc.

GAN with embedded structure



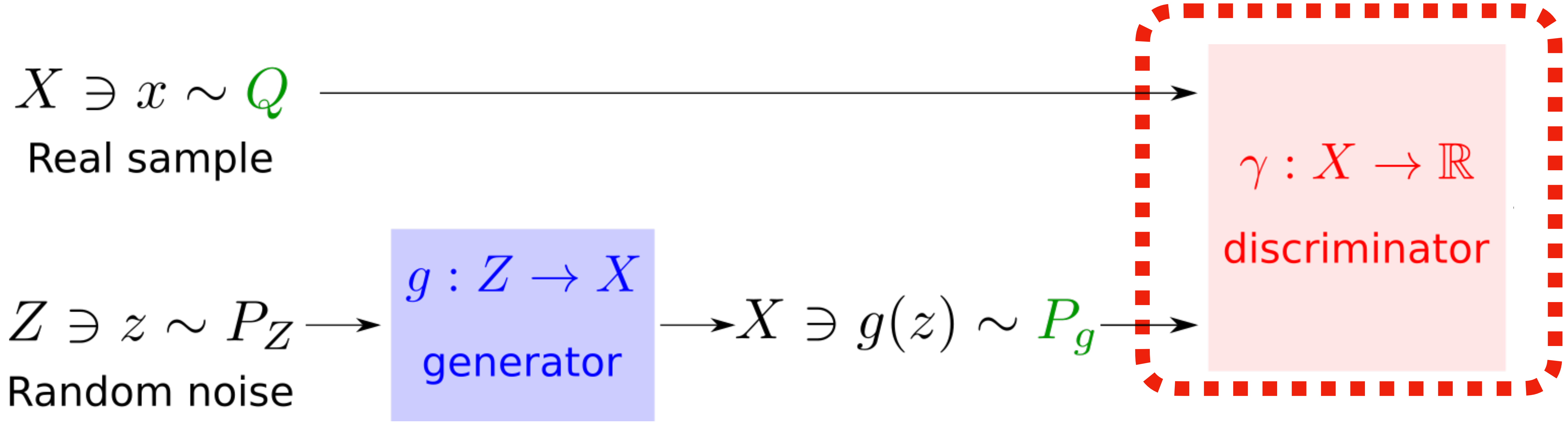
$$\min_{g \in G} D^{\Gamma}(Q \| P_g) = \min_{g \in G} \max_{\gamma \in \Gamma} H(\gamma; Q, P_g),$$

Q is Σ-invariant

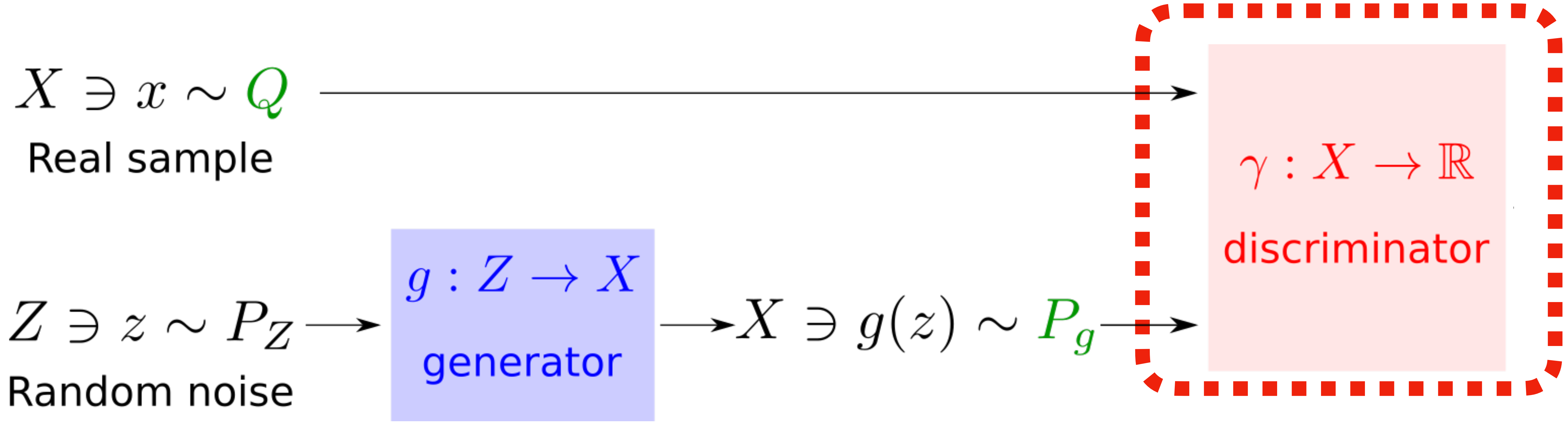


- Target distribution Q is invariant under a group Σ.
- Σ: rotation, translation, roto-reflection, permutation, etc.
- **How to incorporate structure into g and γ?**

Our Theorem 1: “smarter” discriminator



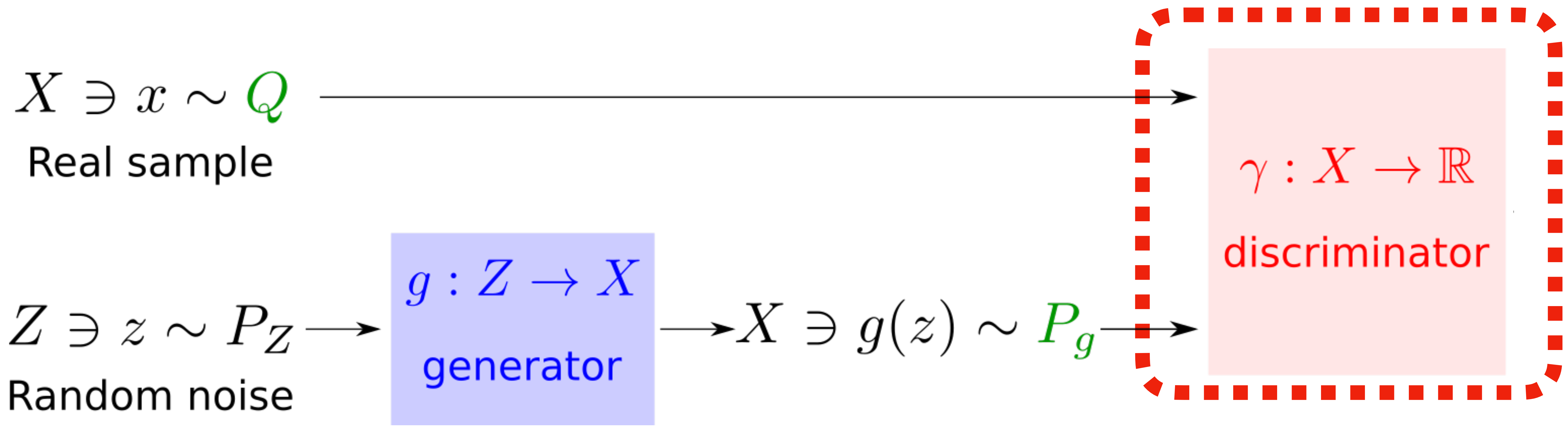
Our Theorem 1: “smarter” discriminator



Theorem: If the distributions P, Q are **Σ-invariant**, then

$$D^{\Gamma}(Q \| P) = D^{\Gamma_{\Sigma}^{\text{inv}}}(Q \| P),$$

Our Theorem 1: "smarter" discriminator

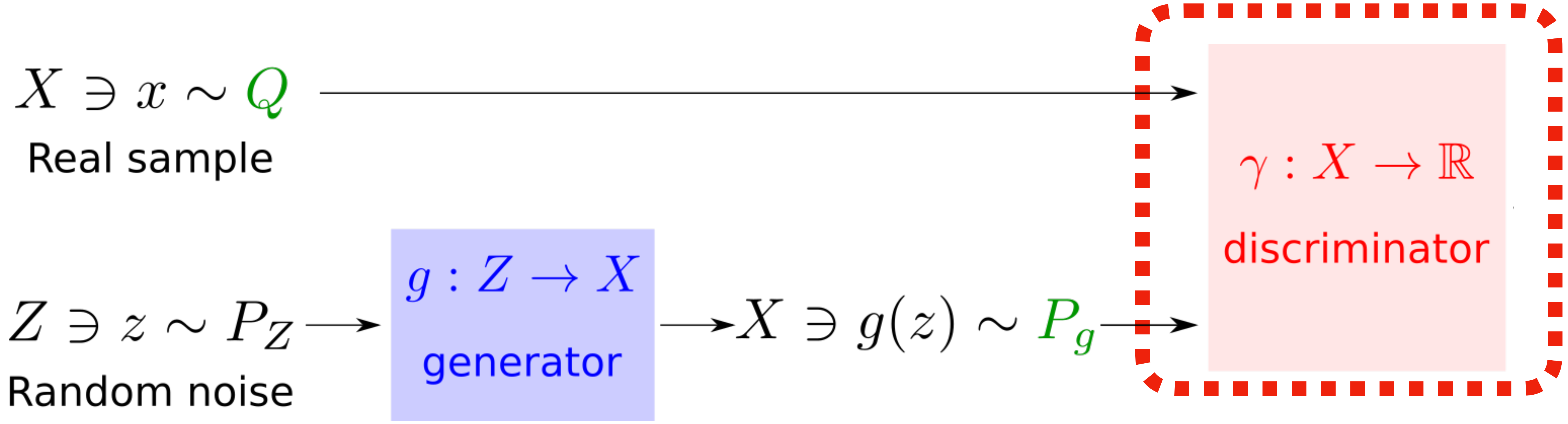


Theorem: If the distributions P, Q are **Σ-invariant**, then

$$D^{\Gamma}(Q \| P) = D^{\Gamma_{\Sigma}^{\text{inv}}}(Q \| P),$$

- Structure information embedded in the "smarter" space $\Gamma_{\Sigma}^{\text{inv}}$ of **Σ-invariant** discriminators

Our Theorem 1: “smarter” discriminator

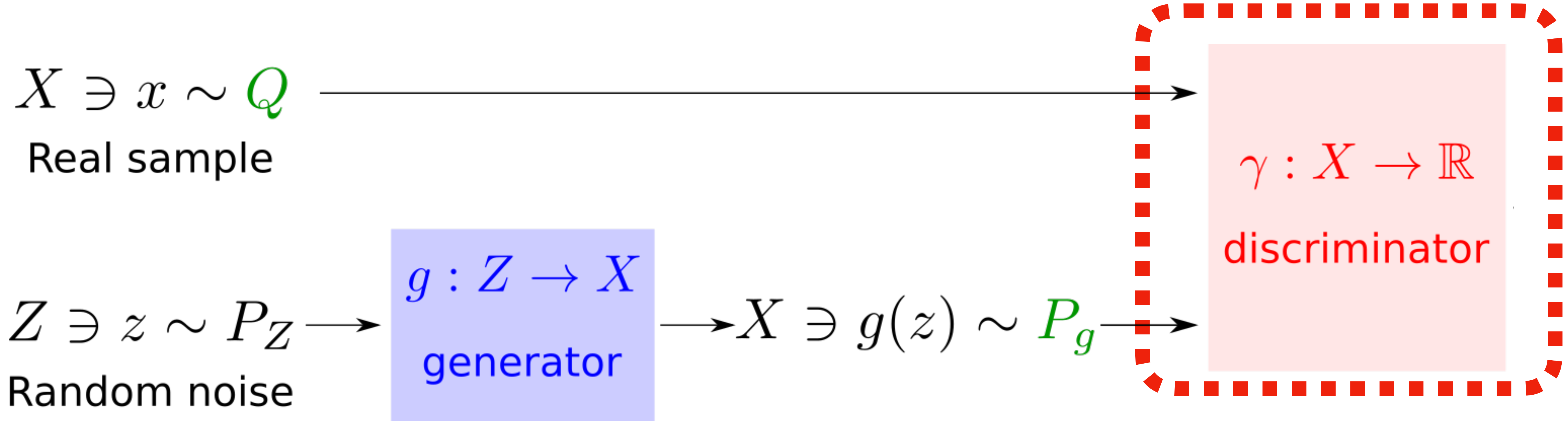


Theorem: If the distributions P, Q are **Σ-invariant**, then

$$D^{\Gamma}(Q\|P) = D^{\Gamma_{\Sigma}^{\text{inv}}}(Q\|P),$$

- Structure information embedded in the “smarter” space $\Gamma_{\Sigma}^{\text{inv}}$ of **Σ-invariant** discriminators
- $\Gamma_{\Sigma}^{\text{inv}}$ is much **“smaller”** than $\Gamma \Longrightarrow$ **efficient** GAN optimization

Our Theorem 1: “smarter” discriminator

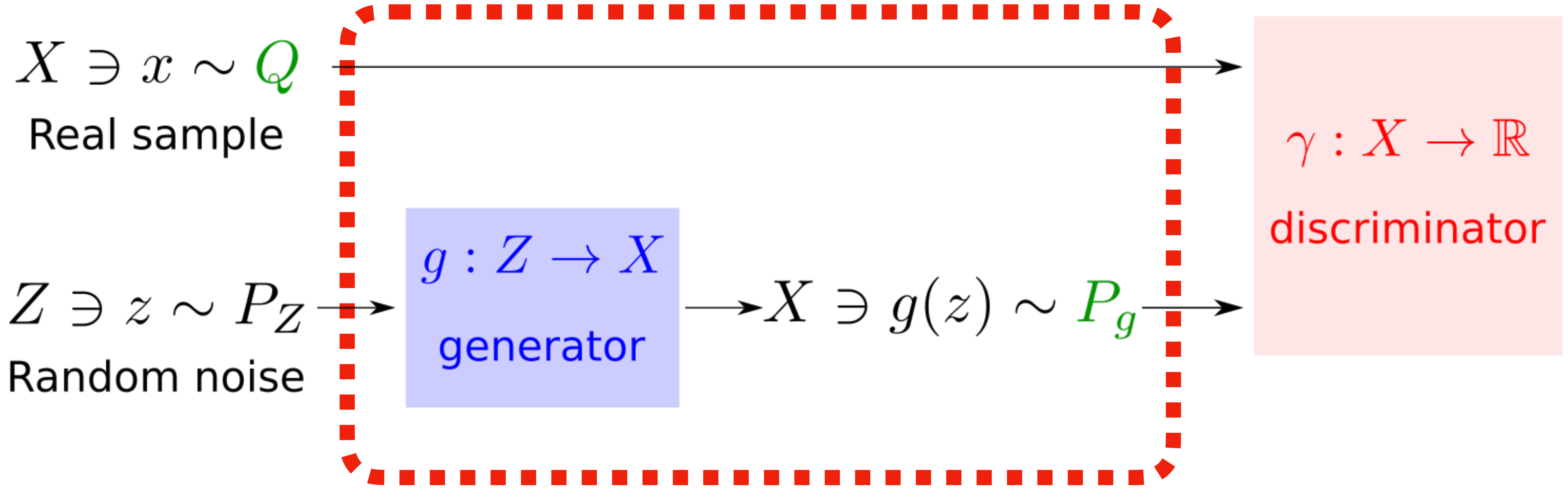


Theorem: If the distributions P, Q are **Σ-invariant**, then

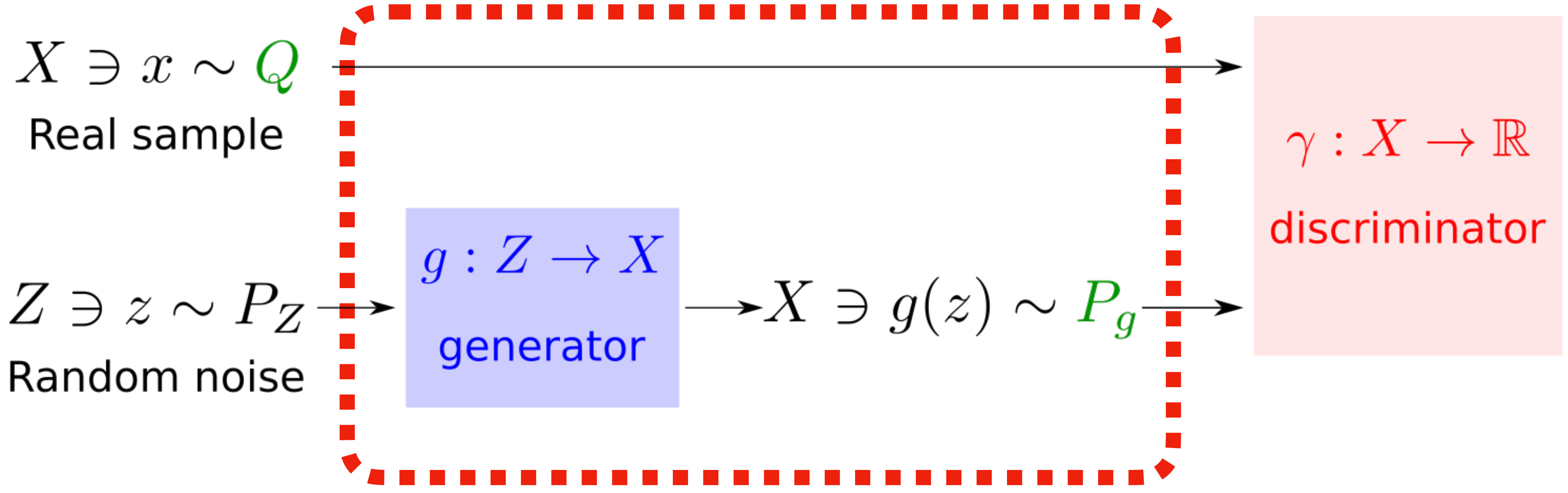
$$D^{\Gamma}(Q \| P) = D^{\Gamma_\Sigma^{\text{inv}}}(Q \| P),$$

- Structure information embedded in the “smarter” space $\Gamma_\Sigma^{\text{inv}}$ of **Σ-invariant** discriminators
- $\Gamma_\Sigma^{\text{inv}}$ is much **“smaller”** than $\Gamma \Longrightarrow$ **efficient** GAN optimization
- $\Gamma_\Sigma^{\text{inv}}$ serves as an **unbiased regularization** to prevent discriminator **overfitting**.

Our Theorem 2: “smarter” generator

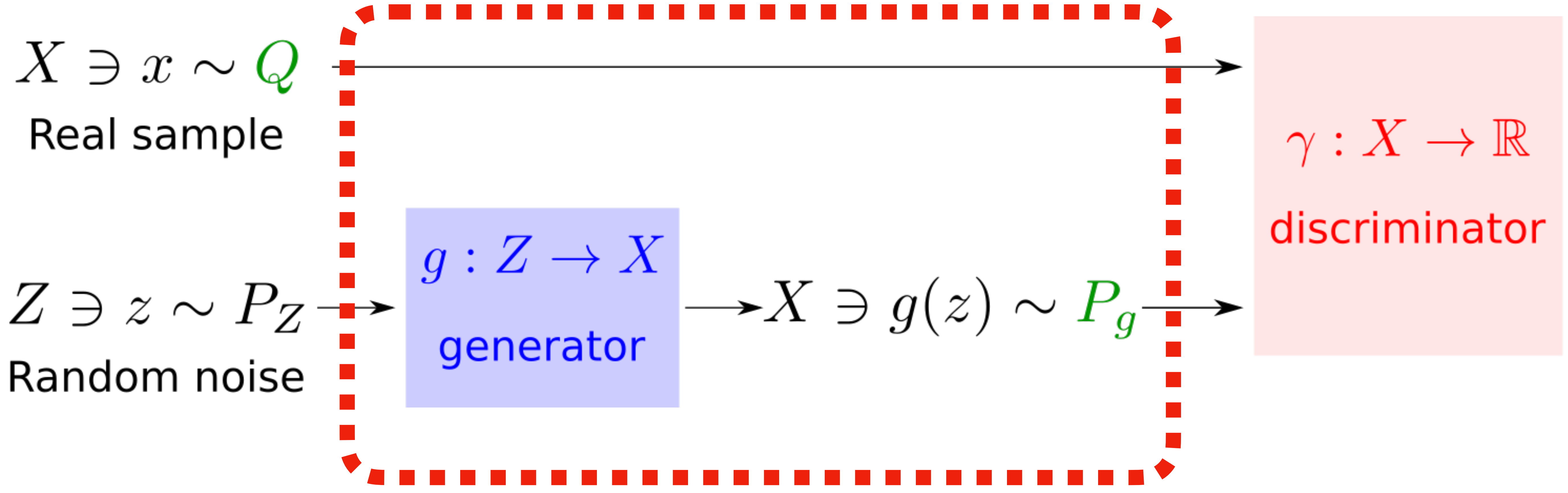


Our Theorem 2: “smarter” generator



Theorem: If P_Z is Σ-invariant and $g : Z \to X$ is Σ-equivariant, the generated measure P_g is Σ-invariant.

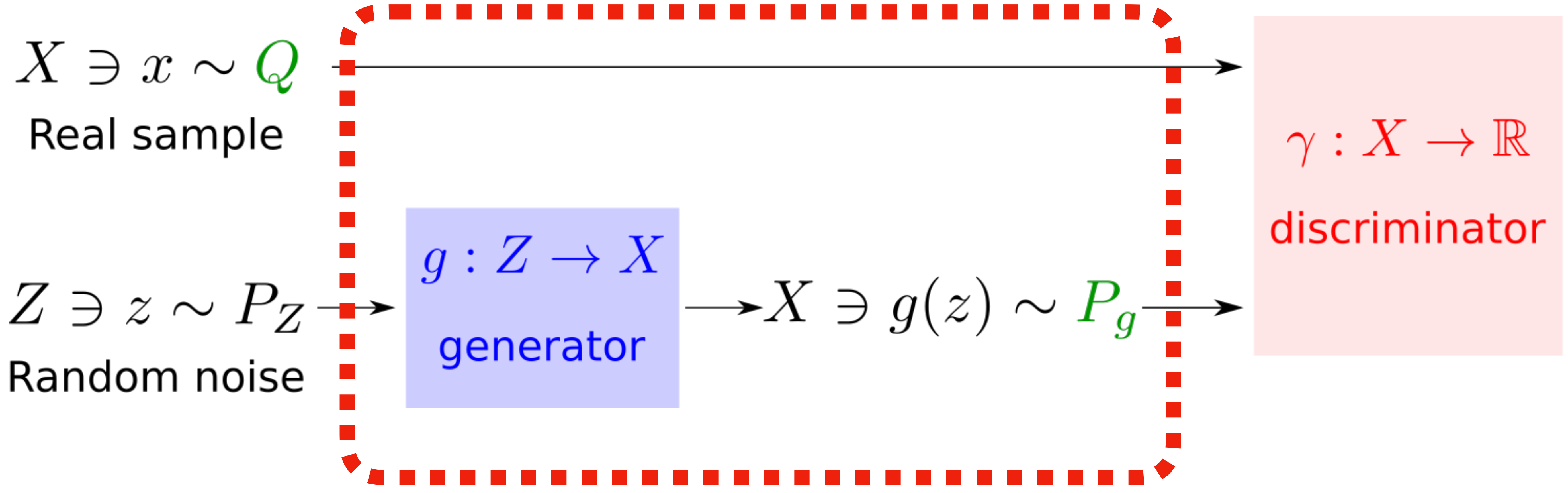
Our Theorem 2: “smarter” generator



Theorem: If P_Z is Σ-invariant and $g : Z \to X$ is Σ-equivariant, the generated measure P_g is Σ-invariant.

- Structure information embedded in the "smarter" generator **and** noise source.

Our Theorem 2: “smarter” generator

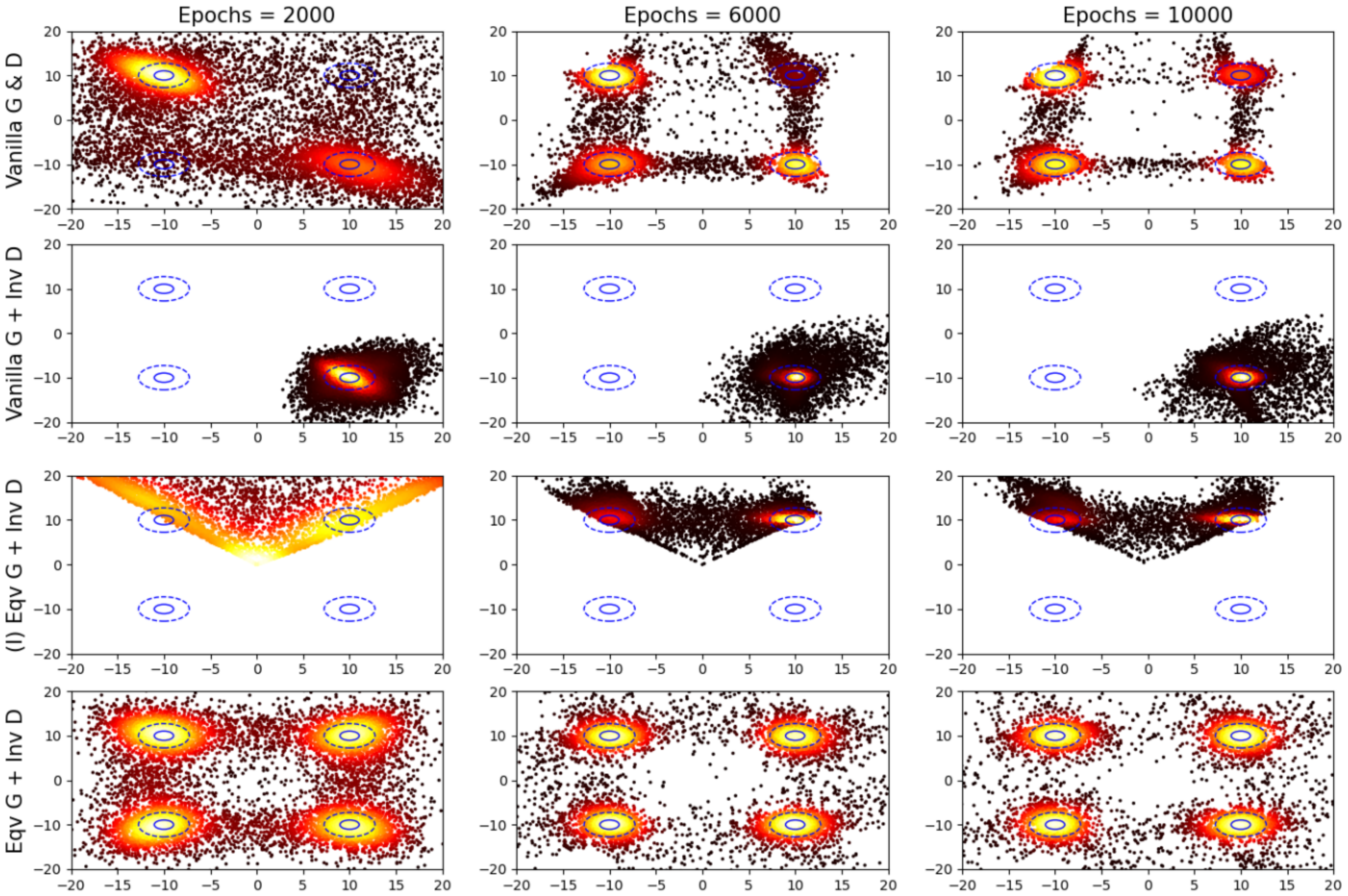


Theorem: If P_Z is Σ-invariant and $g : Z \to X$ is Σ-equivariant, the generated measure P_g is Σ-invariant.

- Structure information embedded in the "smarter" generator **and** noise source.
- "Smart" generator and noise source prevents mode collapse.

Two "smart" players

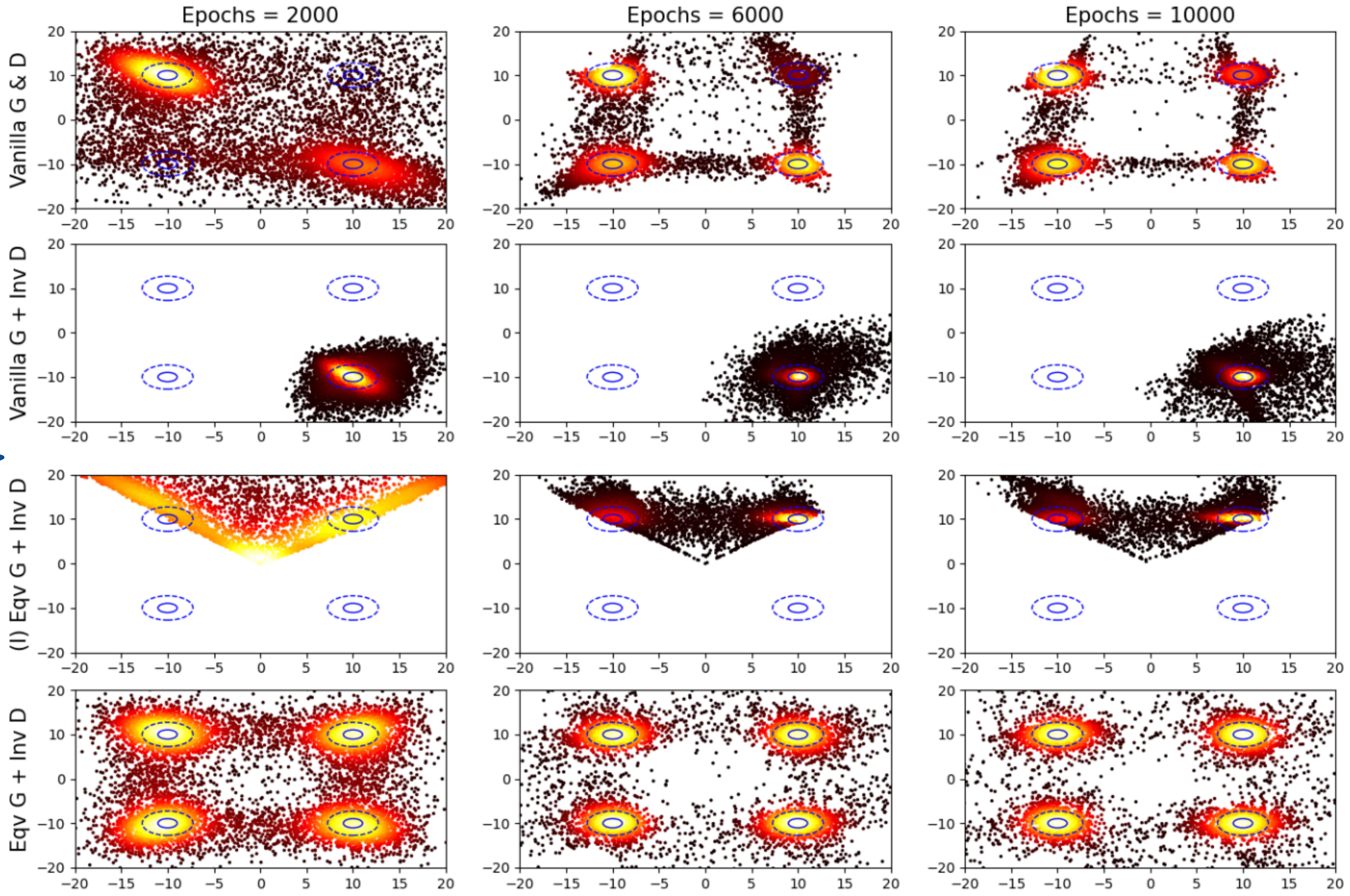
"Ignorant" players need lots of data, lots of time… (the usual GANs)



Two "smart" players

"Ignorant" players need lots of data, lots of time… (the usual GANs)

Players need to be "equally smart": no weak links!

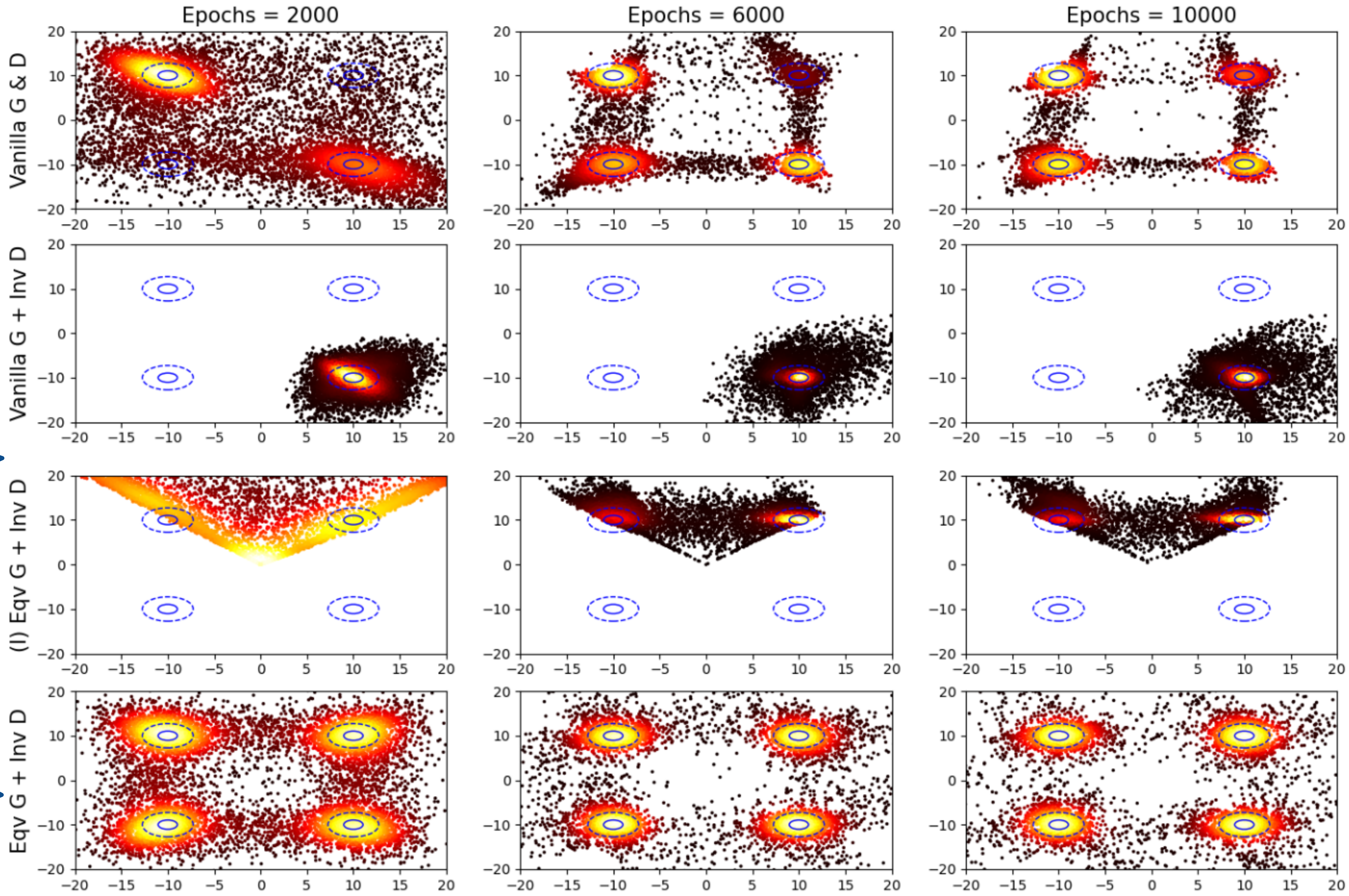


Two “smart” players

“Ignorant” players need lots of data, lots of time… (the usual GANs)

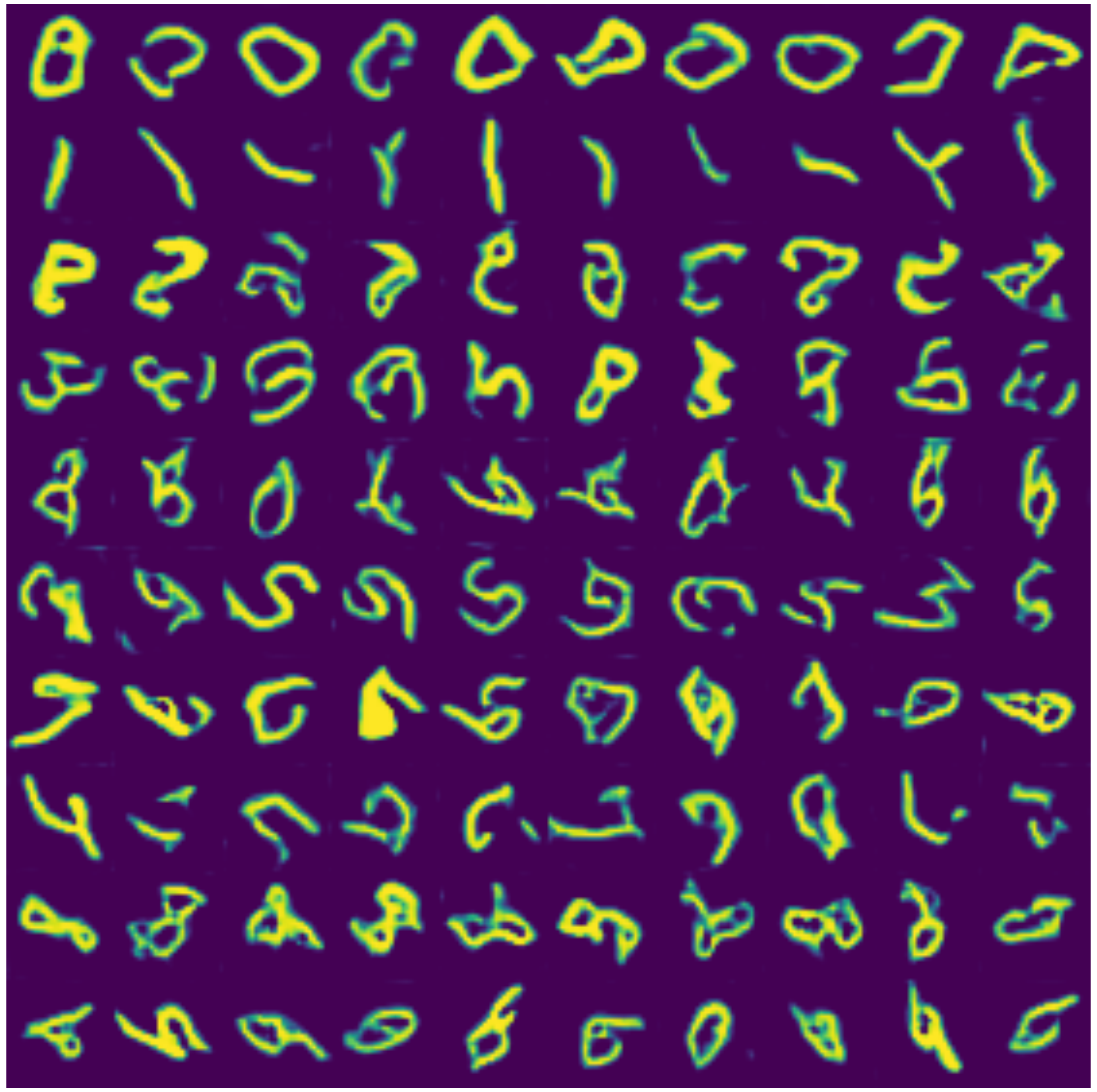
Players need to be “equally smart”: no weak links!

“Smart” players learn faster and better (our GANs)

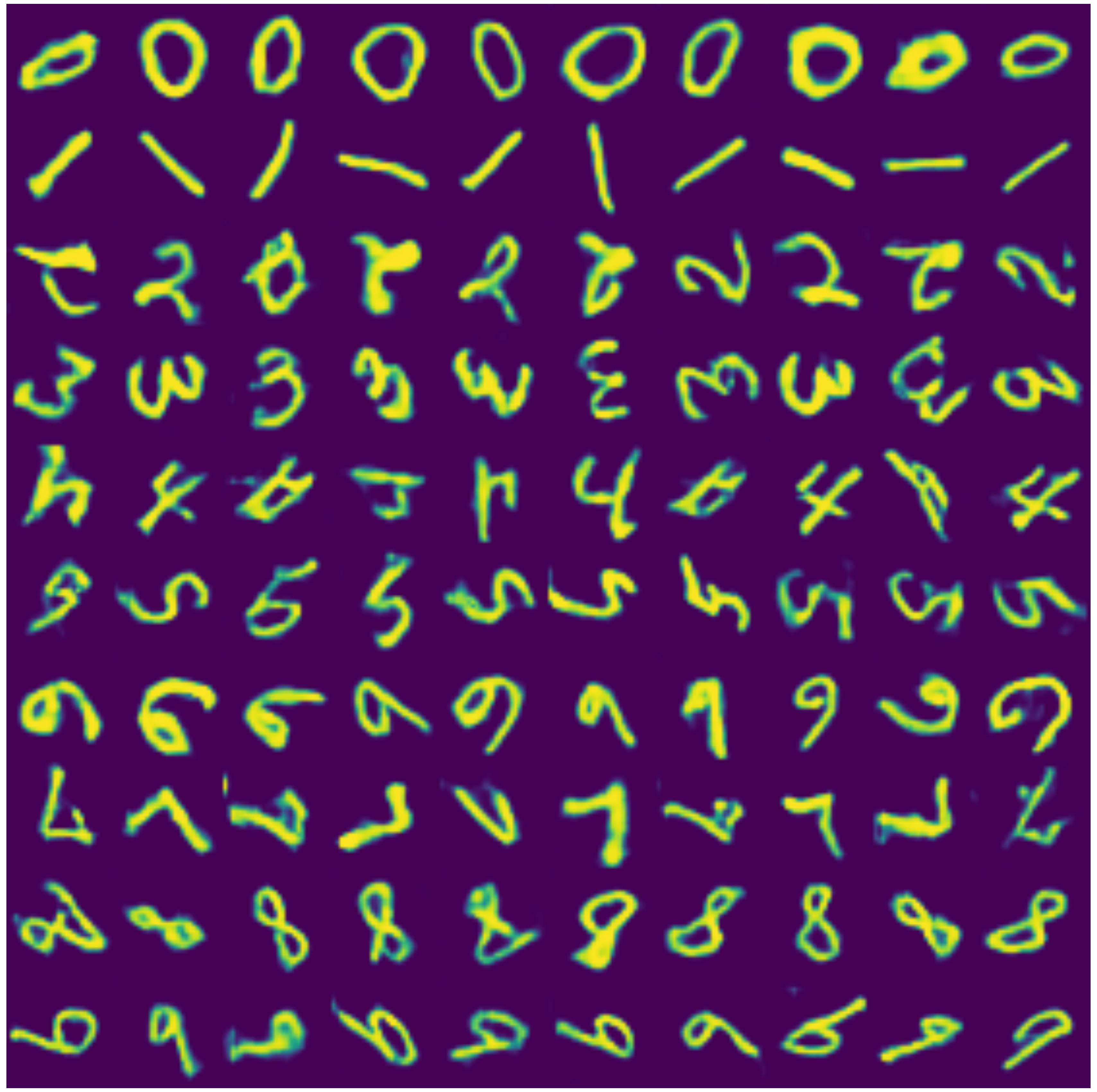


RotMNIST with 1% training samples

“Ignorant” players



“Smart” players

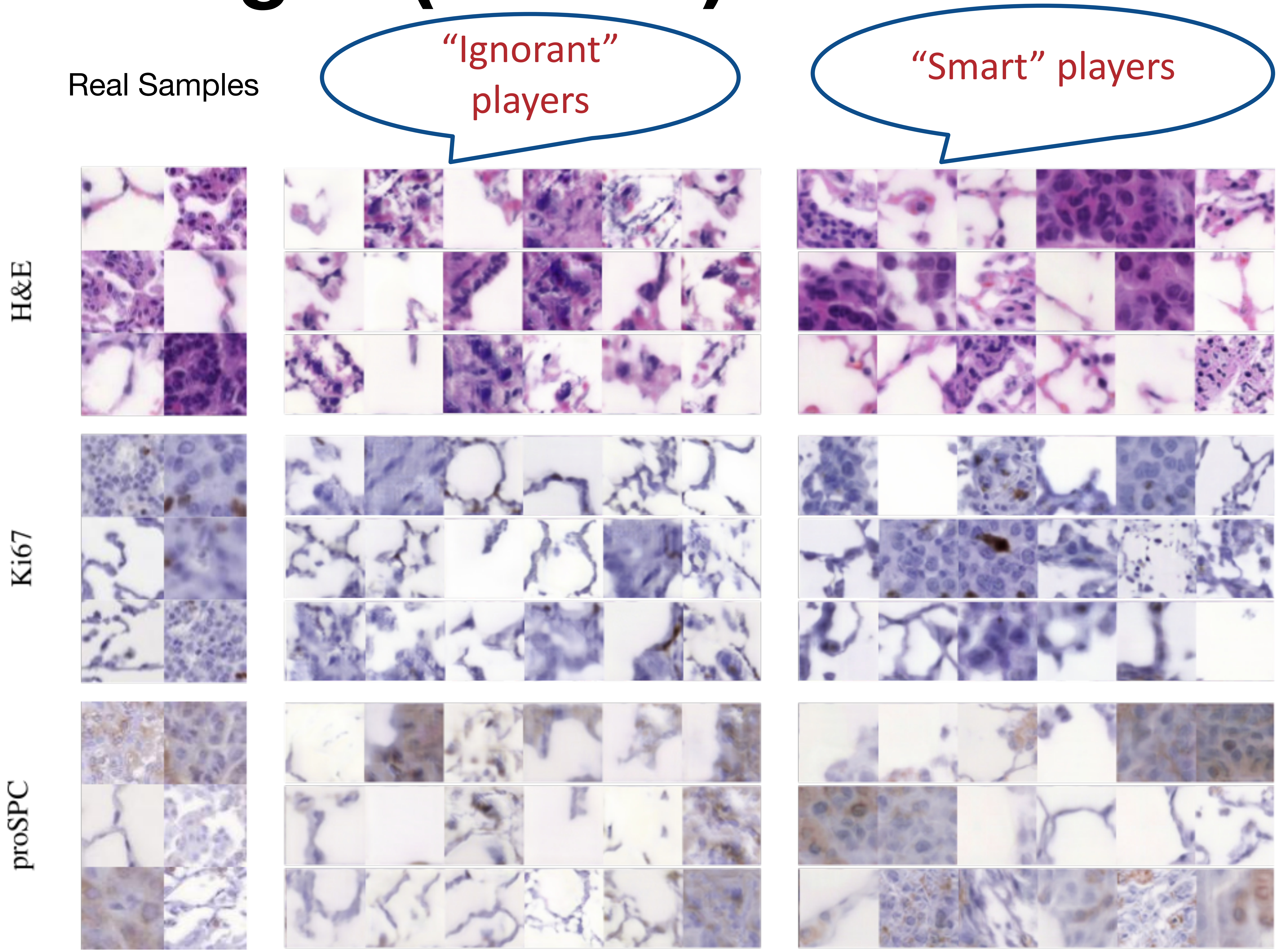


Performance metrics on RotMNIST (FID)

Loss	Architecture	5%	10%	50%	100%
RA-GAN	CNN G&D	357	348	403	392
	Eqv G + CNN D, $\Sigma = C_4$	333	355	380	393
	CNN G + Inv D, $\Sigma = C_4$	181	188	177	176
	(I)Eqv G + Inv D, $\Sigma = C_4$	141	132	135	130
	Eqv G + Inv D, $\Sigma = C_4$	78	89	84	82
	Eqv G + Inv D, $\Sigma = C_8$	**52**	**51**	**52**	**57**
$D^{\Gamma}_{\alpha=2}$-GAN	CNN G&D	261	283	297	293
	Eqv G + CNN D, $\Sigma = C_4$	271	251	274	275
	CNN G + Inv D, $\Sigma = C_4$	208	192	183	173
	(I)Eqv G + Inv D, $\Sigma = C_4$	147	133	124	126
	Eqv G + Inv D, $\Sigma = C_4$	99	88	80	81
	Eqv G + Inv D, $\Sigma = C_8$	**55**	**57**	**53**	**51**

Almost an order of magnitude improvement.

Medical images (ANHIR)



Medical images FID (ANHIR)

Loss	Architecture	ANHIR	ANHIR+
RA	CNN G&D	(186, 523)	(184, 503)
	(I)Eqv G + Inv D	(100, 142)	(88, 140)
	Eqv G + Inv D	**(78, 125)**	**(84, 118)**
D_2^L	CNN G&D	(313, 485)	(347, 539)
	(I)Eqv G + Inv D	(120, 176)	(119, 177)
	Eqv G + Inv D	**(97, 157)**	**(90, 128)**

Loss	Architecture	LYSTO	LYSTO+
RA	CNN G&D	(281, 340)	(250, 312)
	(I)Eqv G + Inv D	(218, 272)	(212, 271)
	Eqv G + Inv D	**(175, 238)**	**(181, 227)**
D_2^L	CNN G&D	(289, 410)	(265, 376)
	(I)Eqv G + Inv D	(253, 343)	(244, 329)
	Eqv G + Inv D	**(205, 259)**	**(192, 259)**